UNIVERSIDADE NOVA DE LISBOA
Faculdade de Ciência e Tecnologia
Departamento de Engenharia Electrotécnica e de Computadores

An High-Speed Parametric ADC and a Co-designed Mixer for CMOS RF Receivers

por
João Manuel Graça Ferreira

Dissertação apresentada na faculdade de Ciências e Tecnologia da
Universidade Nova de Lisboa para a obtenção do grau de
Mestre em Engenharia Electrotécnica e de Computadores

Orientador: João Carlos da Palma Goes
Co-Orientador: João Pedro Abreu de Oliveira

2010

"Pensar simples[1] *. . . "*

A. Steiger Garção

[1]"Keep Simple Thinking".

Acknowledgments

The acknowledgments is the most difficult text that I have to write, mainly because if the reader is paying attention to this lines is because he is expecting to find his name. Others will simply skip this well deserved credits to people I love, respect and esteem.

I would like to start with my Father. His "hard work" mentality and dedication has always been an inspiration, and his wise words guided me in every important steps of my life. Life was tough but he was there for me and my Sister when we needed him. Next, my beloved Sister. My reason to live and stay out-of-trouble during a big part of my life.

The backbone of my closest family is my dearest grandmother. Saturday and holidays family reunions with my Father, Sister, Uncles, Aunts and Cousins at her home kept us close to each other. Having an united family to trust when things go wrong and to enjoy and share happiness, is something that I greatly value.

Thankfully, I am proud to say that I have more good friends than the space to mention them all. So I can't do it. However, João and Celeste, from Sportkart, were pretty much family during the time I worked and raced with them. Their support was not forgotten and this credits are the least that I could do to show my appreciation.

In the academic world I have to start with my friend and work-partner Iuri. During the last few years I have shared success, as well as frustrations, workload and personal issues, with him. I am sure that without this good and trustfully partner and friend things would have been much harder. Also, special word for my advisors, professors João Oliveira and João Goes. Together with professor Luis Oliveira, they created a very challenging environment that guided me trough the state of the art world of electronics. I would like to say that they are much more than just good teachers.

Last, but definitely not least, my better half Patrícia, for keeping me in-track during the ups and downs of these last years. I really "hit the jackpot" when I found her...

The task of acknowledging everyone who I wish and feel that, for some reason, played a important role in my personal or academic life is an impossible mission, so I must restrict this to family and very close work friends. However, if the reader's name is not here, rest assure that I remembered your contribution.

A special thanks to the FCT/MCTES by the financial support through LEADER, SPEED and IMPACT projects.

UNIVERSIDADE NOVA DE LISBOA

Abstract

Faculdade de Ciência e Tecnologia
Departamento de Engenharia Electrotécnica e de Computadores

Mestre em Engenharia Electrotécnica e de Computadores

by João Manuel Graça Ferreira

The rapid growth of wireless communications and the massive use of wireless end-user equipments have created a demand for low-cost, low-power and low-area devices with tight specifications imposed by standards. The advances in CMOS technology allows, nowadays, designers to implement circuits that work at high-frequencies, thus, allowing the complete implementation of RF front ends in a single chip.

In this work, a co-design strategy for the implementation of a fully integrated CMOS receiver for use in the ISM band is presented. The main focus is given to the Mixer and the ADC blocks of the presented architecture.

The traditional approach used in RF design requires 50 Ω matching buffers and networks and AC coupling capacitors between Mixer inputs and LNA and LO outputs. The co-design strategy avoids the use of DC choke inductors for Mixer biasing, because it is possible to use the DC level from the output of the LNA and the LO to provide bias to the Mixer. Moreover, since the entire circuit is in the same chip and the Mixer inputs are transistors gates, we should consider voltage instead of power and avoid the 50 Ω matching networks.

The proposed ADC architecture relies on a 4-bit flash converter. The main goals are to achieve low-power and high sampling frequency. To meet these goals, parametric amplification based on MOS varactors is applied to reduce the offset voltage of the comparators, avoiding the traditional and power-consuming approach of active pre-amplification gain stages.

Contents

List of Figures

List of Tables

Abbreviations

1dB CP	1 dB **C**ompression **P**oint
ADC	**A**nalog-to-**D**igital **C**onverter
BB	**B**ase **B**and
CAD	**C**omputer-**A**ided **D**esign
CMOS	**C**omplementary **M**etal-**O**xide-**S**emiconductor
CG	**C**onversion **G**ain
DAC	**D**igital-to-**A**nalog **C**onverter
DCR	**D**irect-**C**onversion **R**eceiver
DNL	**D**ifferencial **N**on-**L**inearity
DSB	**D**ouble-**S**ide**B**and
ENOB	**E**ffective **N**umber of **B**its
HD$_n$	n^{th}-order **H**armonic **D**istortion
IC	**I**ntegrated **C**ircuits
IIP3	Third-order **I**nput **I**ntercept **P**oint
IF	**I**ntermediate **F**requency
INL	**I**ntegral **N**on-**L**inearity
LNA	**L**ow-**N**oise **A**mplifier
LO	**L**ocal **O**scillator
LPF	**L**ow-**P**ass **F**ilter
LSB	**L**east **S**ignificant **B**it
NF	**N**oise **F**igure
THD	**T**otal **H**armonic **D**istortion
PA	**P**ower **A**mplifier

QAM	**Q**uadrature **A**mplitude **M**odulation
QPSK	**Q**uadrature **P**hase-**S**hift **K**eying
RF	**R**adio **F**requency
S&H	**S**ample **&** **H**old
SC	**S**witch **C**apacitor
SFDR	**S**purious **F**ree and **D**ynamic **R**ange
SNDR	**S**ignal-to-**N**oise and **D**istortion **R**atio
SNR	**S**ignal-to-**N**oise **R**atio
SoC	**S**ystem **o**n **C**hip
SSB	**S**ingle-**S**ide**B**and
WSN	**W**ireless **S**ensor **N**network

In memory of my Dearest Mother. . .

Chapter 1

Introduction

1.1 Background and Motivation

The developments in CMOS technology, verified in the past decade, have allowed the design of circuits operating at very high speed. In recent years, the growing demand for wireless devices turned this low-cost technology in a attractive platform for development of radio frequency (RF) front end circuits. The main goal for designers is to optimize power dissipation in known architectures and/or develop new ones in order to achieve a complete wireless system inside a single chip. This should result in extremely compact devices while maximizing power efficiency.

The main obstacle in wireless communications is the need to deal with a very aggressive medium, where noise and different strong and weak signals must co-habitat without significant interference. This represents added difficulties and trade-offs during the design. As a result, the process to achieve good performance no longer can be seen as isolated circuits, optimized locally, joined together. A co-design approach, optimized for the system, is expected to produce much better results.

The motivation for this work is to apply this co-design methodology to implement a fully integrated receiver. The use of integrated CMOS technology should allow the implementation of every building blocks (except *maybe* the antenna), thus reducing the number of off-chip components.

Due to the nature of this work the main focus is over the Mixer circuit and the analog to digital conversion, and a complete receiver will be left for future work. However, this

two main blocks play an important role in the overall architecture and should allow some considerations and conclusions about the either or not the approach is valid.

1.2 Thesis Organization

The thesis has been organized in five chapters, including this introductory one.

In chapter 2, an RF Receivers overview is made. This overview starts with the basics of communication systems and RF design considerations. Following it, a more detailed description of three possible architectures for receivers are given: Heterodyne, Homodyne and Low-IF (Intermediate Frequency). For each one, known problems and common adopted solutions are discussed with the conclusion that the Low-IF is one of the most appropriate for fully integrated receivers.

Chapter 3 deals with the receiver's analog part, i.e., the Low-Noise Amplifier (LNA), Local Oscillator (LO) and Mixer. The main focus is given to the latter as a center piece in a co-design strategy. Some important guidelines of the dimensioning processes and simulation results are presented. Finally, the last section in this chapter is dedicated to a few conclusions about this subject.

In chapter 4, the interface between the analog and digital domains is dealt. It starts with an overview of the most important blocks in an Analog-to-Digital Converter (ADC) with special emphasis on the Flash architecture. Following it, a theoretical analyses is presented on the Sample&Hold (S&H) circuit, where special attention is given on switches and MOSFET based capacitors with parametric amplification. Next, a parametric based comparator build around a dynamic latch circuit is introduced.

Chapter 5 deals with the complete architecture for the 4-bit Flash ADC. Transistor sizing procedures and simulation results are followed by some conclusions over the ADC's results.

This work ends with a final chapter dedicated to some final conclusions and further research suggestions.

1.3 Contributions

The Parametric Amplification principle used in the 4-bit Flash architecture is a valid alternative to the traditional pre-amplification stage, necessary to reduce the input offset of the comparator, as long as low resolution fulfills the system requirements. The result of the study presented in this work was also applied in an 8-bit Interleaved CMOS Pipeline ADC to realize the required 1.5b and 2b Flash ADC. This circuit was implemented and measured [1].

The co-design strategy applied in the Mixer, LNA and Oscillator in an Low-IF architecture can be a break-through in future receiver's front-ends because it has the potential to achieve circuits with reduced die area, low power consumption and minimum off-chip components. When implemented in CMOS technology, the ADC can be included in the same chip, resulting in a fully integrated RF front-end.

The *Mixer in a Co-Design Strategy* subject presented in this work was submitted to the MIXDES, 16th International Conference (2009), entitled *LNA, Oscillator and Mixer Co-Design for Compact RF-CMOS ISM Receivers* [2], and awarded with an *Outstanding Paper Award* distinction. An extended version, entitled *Co-Design Strategy Approach of LNA, Oscillator, and Mixer* [3] was published in the International Journal of Electronics And Telecommunications Quarterly.

The Pipeline work, entitled *An 8-bit 120-MS/s Interleaved CMOS Pipeline ADC Based on MOS Parametric Amplification [1]* was published in the IEEE TCAS-II, February 2010.

Although the full receiver's architecture has not been simulated or tested, the achieved results lead to believe that a fully integrated receiver with competitive power consumption and area can be implemented.

The main contribution of this work is the personal knowledge acquired as a result of the analyses of other authors work and the feedback from the project advisors.

Chapter 2

RF Receivers Overview

The market interest in portable wireless devices is leading to research into new Integrated Circuit (IC) technologies, circuit configurations and transceiver architectures. Low-power miniature radio transceivers are sought to communicate digital data in cellular telephones, wireless networks and wireless sensors. Many semiconductor technologies are competing today to supply RF-IC's front-ends for different applications, in many cases aiming for low cost solutions with high energy efficiency.

The various design styles and levels of integration are compared, with the conclusion that single-chip silicon transceivers, combined with architectures which substantially reduce off-chip passive components, will continue to dominate digital cellular telephones in the near future [4].

Due to its extensive use in the digital domain, the CMOS technology has been pushed to follow the Moore's law[1]. Consequently, the minimum transistor size has been reduced not only to increase the number of devices *per* unit of area but also to increase its maximum operation frequency (f_T).

Therefore, the CMOS technology is, nowadays, one of the most attractive choices to implement mixed-mode circuits, reducing manufacturing costs. Developments in this technology lead to believe that it is able to achieve high performance at high frequencies, replacing other expensive technologies, and allowing designers to merge all circuits in one single chip (SoC - System on Chip), ideally without requiring any external components.

[1]Moore's laws states the transistor density of integrated circuits doubles every 2 years.

Following sections of this chapter will briefly discuss conventional architectures for receivers, namely, *Heterodyne*[2], *Homodyne* or *Zero-IF* and *Low-IF*. Some advantages and disadvantages for each one are appointed and an appropriated one is selected.

The next two chapters will focus on two building blocks, Mixer and ADC, that are part of the receiver's architecture. The chosen topology for each one is presented and discussed.

2.1 Introduction

The basic configuration of a wireless communication system is presented in figure 2.1. The main goal of a communication system is to deliver some kind of information from one system to another.

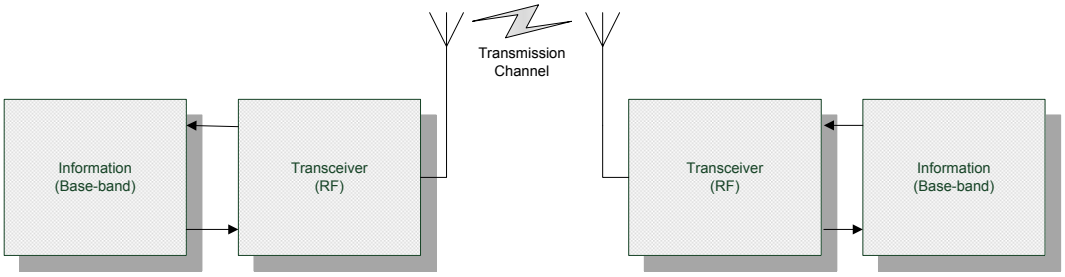


FIGURE 2.1: Basic configuration of a wireless communication system

The information on the sender side, disregarding the method used to gather it, is usually an electric signal located in a baseband region of the frequency spectrum. This signal is then delivered to a transmitter, which is responsible for transfer the baseband signal to an higher frequency. This operation is important because low frequencies require longer antennas and carry fewer information than high frequencies (however, high frequencies suffer from attenuation). The translated signal is then amplified by an amplifier commonly known as a Power Amplifier (PA), and applied to an antenna that radiates an electromagnetic field through the radio channel (air, vacuum, etc.).

In the receiver side, the reverse operation is performed, i.e., the antenna is excited by an electromagnetic field and produces an electric signal. Due to the hostile nature of the transmission channel, the signal received is usually very weak and noisy. To processes this weak signal, a special type of amplifier (the LNA) is used to amplify it without adding any noise (ideally). After the amplification, the receiver performs a frequency translation

[2]Heterodyne derives from *hetero* (different) and *dyne* (to mix).

to a lower frequency (down-conversion) in order to process the signal and retrieve the original information in the base-band.

There are several architectures for transceivers, which include both the receiver and the transmitter in the same device using the same antenna, depending on the type of information (digital or analog), power constraints, standard technologies' specifications, etc. In this work we will focus only on receivers and it's building blocks, giving special attention to the Mixer and the ADC, responsible for the down-conversion mentioned above and the transformation from the analog to the digital domain, respectively.

2.2 RF Systems

RF systems differ from other baseband analog or mixed-signal systems because high frequency and transmission line effects have to be taken into account. High frequencies make small parasitic capacitances and inductances considerable, so disregarding them is no longer valid and proper modeling of devices is a key-factor. Also, as the frequency increases, the wavelength decreases and, consequently, the circuit boards' tracks have lengths similar to the signal wavelength. This means that the wire or track that connects two devices, no longer behaves as a short-circuit and power transfer between blocks must be carefully analyzed.

Another important subject is related to the hostile transmission channel, shared by different devices, meaning that the signal of interest is surrounded by others and, sometimes, destructive overlaps occurs. So the receiver must be able to detect the proper signal and reject all other, by means of filtering, time division schemes, channel separation and others, beyond the scope of this work [5]. Nowadays, communications standards define, among others, frequencies to use (actually, spectrum usage is a legal issue), channel usage, modulation schemes, encryption, registration process, data type, while mobility (or battery life vs. battery size) defines power constraints.

It is clear that an RF designer must have knowledge in a variety of fields, such as signal propagation, wireless standards, transceiver architectures, random signals, integrated circuit design, CAD tools, etc., making it difficult for a single designer to have profound knowledge in all subjects. This leads, in many cases, to commercial applications that result from separated developed blocks, glued together. This has proved to be inefficient and redundant and RF designers appear, to implement fully integrated solutions.

2.3 Receiver's Architectures

In order to properly decide which is the best[3] architecture for a certain application, some knowledge about the most widely used architectures is required. Studying these architectures also provide useful information about the problems that exist in wireless communication environments.

The wireless communications environment is many times referred as *hostile* because it imposes severe constraints upon transceivers design. Limited spectrum allocated for each user, weak signal reception "mixed" with strong interferer signals and multipath are some of the problems that have to be solved under low power constraints, especially in portable devices.

The Heterodyne architecture is still widely used in wireless systems, mainly because of its high performance. It was first proposed by Edwin Armstrong in 1917 and has been the dominant choice in RF systems until recently. Although, the need to integrate RF front-end in CMOS technology, has pushed the use of both Homodyne and Low-IF topologies.

2.3.1 IF or Heterodyne Receivers

Wireless standards define a limited spectrum for each user. For example, the GSM-900 standard uses the 890-915 MHz band for uplink and 935-960 MHz for downlink. Each of them is divided in 124 channels with 200 KHz bandwidth. A receiver for this system must be able to handle signals in the 935-960 MHz band and, within this band, distinguish each channel of 200 MHz.

Figure 2.2 clarifies the difference between channel and band. It also shows that signals from other standards may (and certainly will) exist outside the allocated band. This out of band signals can be rejected by an wide bandpass filter, covering the entire band of operation, named *Band Select filter*.

The narrow bandwidth available to each channel means that the receiver must be able to process the desired one and reject all adjacent channels. For this reason, a second bandpass filter is used. This one is commonly called *Channel Select filter*

[3]It should be kept in mind that there is not an overall *best* (optimum) solution. Decisions are usually based on trade-offs between different performance properties (linearity, gain, power consumption, die area, etc.)

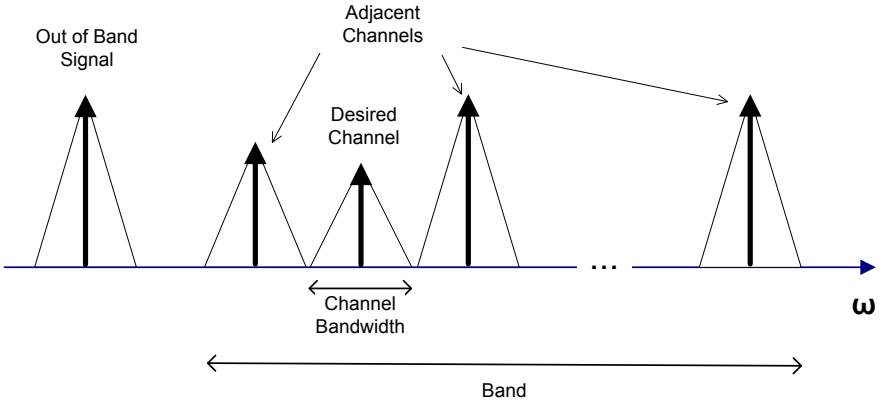


FIGURE 2.2: Wireless spectrum.

However, the realization of such filters is a major bottleneck. Selection of a channel with high center frequency closely surrounded by others, require high quality factor (Q) for this filters, which are difficult to achieve even with external surface acoustic wave (SAW) devices ([5]) and impossible to realize in SoCs approaches.

The major improvement given by the Heterodyne architecture is a solution for the above mentioned problem. The idea is to "transfer" an high frequency signal to an intermediate lower frequency and relax the channel selection filter's specifications. As it will be discussed next, multiplying two signals achieves this "transfer" (or translation) in frequency. This operation is preformed by a *Mixer*.

Consider the ideal situation of a clean spectrum with a single tone at ω_{rf} and amplitude V_{rf} ($v_{rf}(t) = V_{rf} \cdot cos(\omega_{rf} t)$). By multiplying this signal with another single tone at ω_{lo} and amplitude V_{lo} ($v_{lo}(t) = V_{lo} \cdot cos(\omega_{lo} t)$) the following result is obtained:

$$v_{rf}(t) \cdot v_{lo}(t) = \frac{V_{rf} \cdot V_{lo}}{2} \left[cos\left(\left(\omega_{rf} + \omega_{lo}\right) t\right) + cos\left(\left(\omega_{rf} - \omega_{lo}\right) t\right)\right] \tag{2.1}$$

Equation 2.1 shows that this operation produces two signals. One is thrown to an high frequency ($\omega_{rf} + \omega_{lo}$), is not relevant for downconversion and is easily removed by filtering. The other comes to $\omega_{rf} - \omega_{lo}$, which will be called intermediate frequency ω_{if}. This result is better understood in the frequency domain, shown in figure 2.3.

Upon multiplication, with the desired signal present at ω_{if}, the realization of a channel selection filter is much easier. This is more so as ω_{if} decreases. Moreover, changing ω_{lo} between specific values causes that a specific channel can be centered in ω_{if}. Thus, a

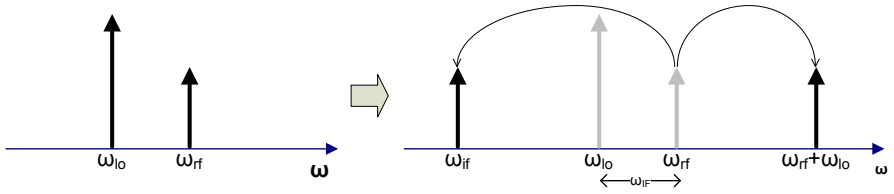


FIGURE 2.3: Multiplication of two tones in frequency domain.

single filter tunned to a static center frequency and bandwidth, can be used for channel selection, making it much more efficient.

The main problem in the previous approach is the assumption of clean spectrum. Consider now that a second unwanted tone, ω_{im}, exists inside the band of operation (thus, not filtered by the band selection filter), and keep in mind that the channel select filter have to be placed *after* the Mixer, due to the above mentioned problems.

Since the multiplication does not preserve the *polarity* of the difference between its two inputs, the output form $cos\,(\omega_{im} - \omega_{lo})\,t$ is no different from $cos\,(\omega_{lo} - \omega_{im})\,t$. Thus, the tones symmetrically located above and below ω_{lo} are downconverted to the same frequency ω_{if} [5], as illustrated in figure 2.4.

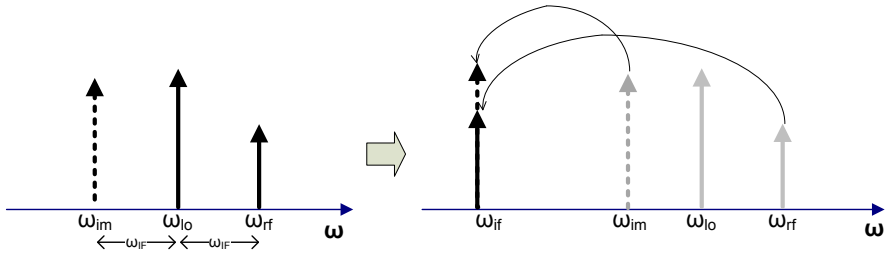


FIGURE 2.4: Problem of image in Heterodyne receivers.

Wireless' standards imposes constraints upon signal emissions in their own users, but may have no control over signal present in other bands[4]. So, it is possible that much stronger signals appear at ω_{if} that will interferer with the desired signal. The traditional approach to solve this problem is to apply a filter, commonly called *Image Reject Filter*, placed before the Mixer, in order to suppress this interferers. Note that in this case, the signal that has to be rejected is $2\omega_{if}$ apart from the desired one. Thus, higher ω_{if} results in loosen specifications for the filter (which is the opposite of what was needed for the channel select filter).

[4]The single tone approach can be easily extended for a channel/band

Another important aspect is the received signal's strength (or, better yet, weakness). Usually, the signal provided by the antenna is very weak. So, some kind of amplification is required. Note that the downconverted signal is proportional to V_{lo} (as shown in equation 2.1), which means that some amplification is possible by increasing this signal's amplitude. However, this usually does not suffice the requirements and an LNA is used to amplify the weak signal without adding too much noise. This circuit is usually placed very close to the antenna to avoid losses and extra noise.

Figure 2.5 shows the complete Heterodyne architecture used in receivers. Note the shaded area that produces a second translation to another IF or baseband (BB). This operation is used to minimize the trade-off required in this architecture. Nowadays, this is usually preformed in the digital domain, i.e., the ADC performs analog to digital conversion at an higher frequency. Additionally, some trade-offs are required: power consumption and area *vs.* ADC performance at high frequency.

The following blocks (such as an DSP - Digital Signal Processor) process the information and retrieve it in baseband. In this case the shaded area is simply removed from the architecture and it is sometimes called *Digital-IF Architecture*.

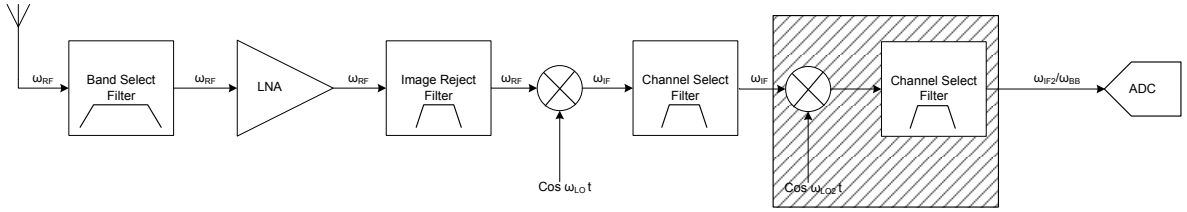


FIGURE 2.5: Heterodyne architecture.

The widely used Heterodyne architectures allows the realization of wireless receivers for standards working at high frequency. It avoids unfeasible channel select filters by introducing an Intermediate Frequency. Choosing this frequency is a critical trade-off in the Heterodyne receiver: with high IF image rejection is easier, whereas with low IF the suppression of adjacent channels is easier.

The image-reject architectures proposed by Hartley and Weaver [6], avoid this trade-off by processing the signal and the image differently. In theory they should work but due to mismatches they have their own limitations or require impractical requirements in other circuits [7].

2.3.2 Zero-IF or Homodyne Receivers

In practice, the heterodyne architecture described above requires the use of an external filter to realize image rejection [5]. This adds some more trade-offs to the circuit due to the requirement of input/output impedance matching (usually to 50 Ω) between LNA-Filter-Mixer. Moreover, the "optimal" IF that results from the trade-off between image rejection and channel selection is still very high and makes it difficult for the ADC to fulfill specifications (specially, in terms of low power consumption and low die area), when the Digital-IF Architecture is used.

This is not a good solution for low-cost, low area, and ultra compact modern applications. The challenge nowadays is to obtain a fully integrated receiver, on a single chip.

Homodyne receivers overcome these difficulties by directly translating the signal to base-band[5]. In this type of receivers the signal's carrier (or center frequency) is the same as the LO's frequency, leading to $\omega_{if} = \omega_{rf} - \omega_{lo} = 0$. As a result, this architecture has three major advantages over the Heterodyne:

- there is no image signal to be rejected;
- channel selection is preformed by Low-Pass Filter (LPF);
- post-processing is at baseband, relaxing requirements for filters and ADCs.

Before discussion about the first advantage, it is important to remember an important fact about the frequency spectrum that has been neglected so far. Until now, only positive values of ω has been considered, but the frequency spectrum also have a negative side, which is a *mirrored* image of the positive one in respect to zero, as illustrated in figure 2.6.

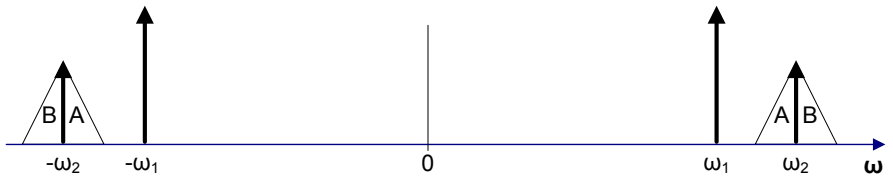
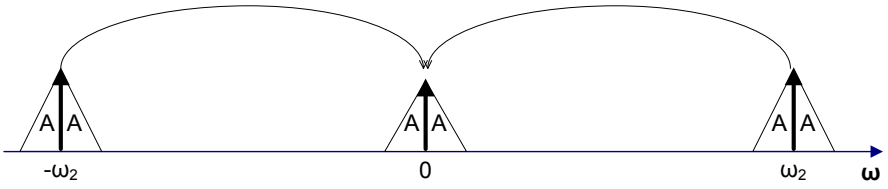


FIGURE 2.6: Negative and positive sides of the frequency spectrum.

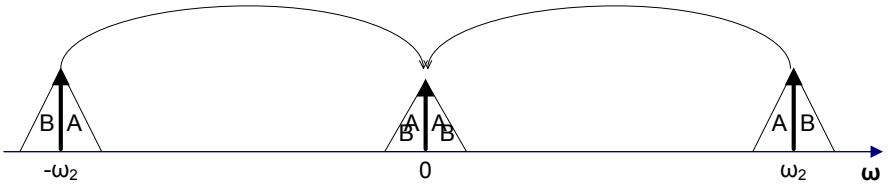
[5]That's why they are also called **D**irect-**C**onversion **R**eceivers (DRC)

As a result, in zero-IF architecture, the downconverted signal and its image are the same, thus there is no need for image rejection.

However, since there is a mirror effect, the translated spectrum will be the superposition of the **A** and **B** parts indicated in figure 2.6. This is not important if both sides are equal (**A** and **B** carries the same information). For example, Amplitude Modulation (AM) produces a symmetric spectrum with respect to the carrier's frequency. These types of signals are usually referred as Double-SideBand (DSB) signals. Note that, since both sides carry the same information, only half of the bandwidth required carries really useful information. Figure 2.7(a) shows a DSB signal downconversion.



(a) DSB downconversion.



(b) SSB downconversion.

FIGURE 2.7: Direct conversion in SSB and DSB signals

Other modulation schemes, such as frequency or phase modulation, make use of all the bandwidth. This means that **A** and **B** parts are different and the de-modulated signal is corrupted by the superposition, as illustrated in figure 2.7(b). This type of signals are usually referred as Single-SideBand (SSB) signals.

For this type of modulation, the downconversion must provide quadrature outputs to avoid loss of information. This quadrature modulation is also referred to as I/Q modulation.

One of the simplest forms of I/Q modulation is the Quadrature phase-shift keying (QPSK) and is briefly discussed next in order to understand how this requirement affects performance.

Consider that the received signal is in the form $v_{rf} = a\,cos\,\omega_{rf}t + b\,sin\,\omega_{rf}t$, where a and b are either $+1$ or -1 (thus, producing a 180° phase shift). Mixing this signal with one provided by an LO with the same frequency, $v_{lo_I} = 2\,cos\,\omega_{lo}t$ (with $\omega_{rf} = \omega_{lo}$) the resulting baseband signal[6] is as follows:

$$v_{BB_I} = a\,cos\,(\omega_{rf} - \omega_{lo})\,t = a \tag{2.2}$$

If the LO produces both the in-phase and a 90°-phase shifted signal (in quadrature), $v_{lo_Q} = 2\,sin\,\omega_{lo}t$ then the following result is achieved:

$$v_{BB_Q} = b\,cos\,(\omega_{rf} - \omega_{lo})\,t = b \tag{2.3}$$

By plotting all possible combinations between $[a, b]$ the obtained I/Q constellation is the one presented in figure 2.8.

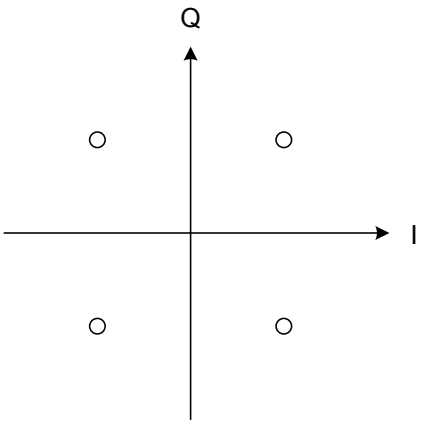


FIGURE 2.8: QPSK ideal constellation.

Note that the received signal v_{rf}, located at ω_{rf}, actually carries two different values: a and b. Reducing those values to $+1$ and -1 (binary system) lead to *2-bits* of information for each received *symbol*. Other modulation schemes, such as Quadrature Amplitude Modulation (QAM), apply the same principle but, instead of a binary system, use ternary, quaternary, in order to achieve more bits *per* symbol (16-QAM, 64-QAM, 128-QAM and 256-QAM, are common forms for this modulation). Naturally, the ideal case is to get as many bits per symbol as possible but the "quality" of the received signal dictates the choice. This will be clarified while discussing some disadvantages in the following text.

[6]The $\omega_{rf} + \omega_{lo}$ result is suppressed because it can be easily removed by filtering.

Figure 2.9 shows the Homodyne architecture with the I and Q signal paths required for proper SSB signals demodulation.

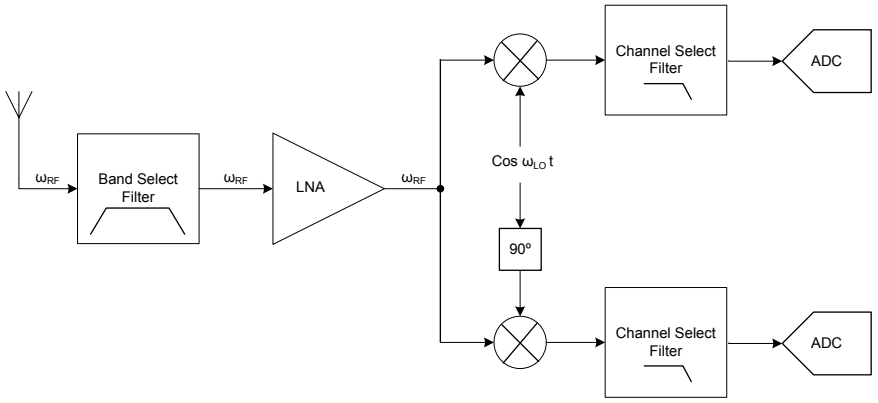


FIGURE 2.9: Homodyne architecture.

The second and third advantages are somewhat obvious. Since the signal is located at baseband, the channel select filter can be realized by a low-pass filter with a cut-off frequency equals to half the bandwidth, which is more simple to implement than a band-pass filter at higher frequencies, and in nowadays technologies does not require external devices. Moreover, analog-to-digital conversion is done at lower frequency, relaxing ADCs requirements.

As a rule of thumb, one must be aware that a solution to a problem creates different ones, and the Homodyne receiver is no exception. Some of the more relevant disadvantages are discussed next:

- Quadrature error - Quadrature error and mismatches between the amplitudes of the I and Q signals corrupt the downconverted signal constellation, shown in figure 2.8. In practice, the received constellation is formed by a scattered points around the ideal one. In noisy environments, the dispersion can lead to a wrong decision by the receiver.

- DC-offset - Since the downconverted band extends down to zero frequency, any offset voltage can corrupt the signal and saturate the receiver's baseband output stages. Hence, DC offset removal or cancellation is required in directconversion receivers.

- LO leakage - LO signal coupled to the antenna will be radiated again and re-injected to the Mixer through the main signal path, originating unwanted baseband DC components.

- Flicker noise - Having a $1/f$ corner at low frequency, this noise can corrupt substantially the low frequency baseband signals, which is a severe problem in MOS implementation.

The direct conversion approach requires very linear LNAs and Mixers, high frequency LOs with precise quadrature, and use of a method for achieving sub-μV offset and $1/f$ noise. All these requirements are difficult to fulfill simultaneously [6].

2.3.3 Low-IF Receivers

It is well known that Heterodyne receivers have important limitations due to the use of external image reject filters. DCR have some drawbacks because the signal is translated directly to the baseband, e.g., the flicker noise is associated with the nature of CMOS device, so there is no outstanding solution to decrease it so far. Thus, there is interest in the development of new techniques to reject the image without using filters. An architecture, which combines the advantages of both the IF and the zero-IF receivers, is the low-IF architecture.

The low-IF receiver is a Heterodyne receiver that uses special mixing circuits that cancel the image frequency, as shown in 2.10. A high quality image reject filter is not necessary anymore, while the disadvantages of the zero-IF receiver are avoided.

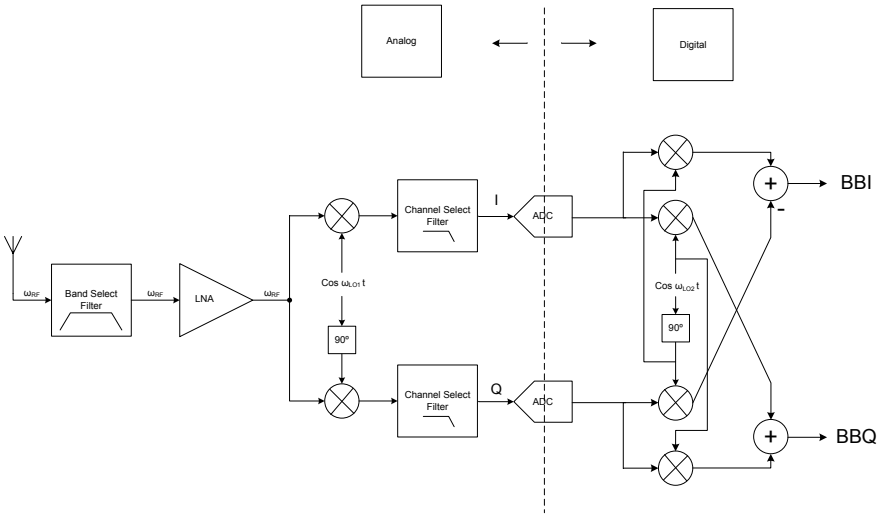


FIGURE 2.10: Low-IF receiver.

As a result, the Low-IF architecture seems to be the obvious choice for fully integrated receivers. The Heterodyne receiver can achieve better performance but is impossible to be fully integrated, whereas the DCR is limited in CMOS technology by the 1/f noise.

2.4 Applications

Wireless Sensor Networks (WSNs) are potential wireless network applications for the following future ubiquitous computing system[7]. Ubiquitous sensor networks are an emerging research area with potential applications in environmental monitoring, surveillance, military, health, and security [8]. The power dissipation of WSNs is critical because this devices require low power consumption for several years operation.

There has been a great deal of interest in realizing low power, low cost, compact RF-IC transceivers for WSNs. Several technological trends that are driving the technical evolution of wireless technology include the process scaling of CMOS transistors and higher bandwidth available at industrial, scientific, and medical (ISM) bands. Almost all of the license free bands propose both linear and nonlinear modulation standards for wireless applications, thus, requiring different *design optimizations* in the RF transceiver. Along with these issues, there exists the challenge to develop *fully integrated* wireless solutions in silicon-based substrates [9].

The communication nodes for ubiquitous networks are required to be integrated in one die for low power consumption and low cost wireless sensor network applications. The overall wireless personal area network (WPAN) system architecture presented in [8] consists of the RF transceiver and a companion digital baseband processor, which implements both physical (PHY) and medium access control (MAC) layers of the IEEE 802.15.14 standard, in sub-gigahertz band ($900MHz$). In this work ([8]), the RF transceiver chip includes a 6-bit digital-to-analog converter (DAC) for the transmitter, and a *4-bit I/Q analog-to-digital converter* for the receiver.

[7]Ubiquitous computing is a post-desktop model of human-computer interaction in which information processing has been thoroughly integrated into everyday objects and activities.

Chapter 3

Mixer in a Co-Design Strategy

In this section, a co-design strategy for the implementation of a low-voltage fully integrated CMOS receiver is presented. With the increase of the transistor's cutoff frequencies, both the DCR and the low-IF receiver techniques allow significant reduction of the number of off-chip components, which means that all the major building blocks will interconnect to each other inside the chip. Therefore, the match between these internal interconnects at 50 Ω level is no longer required. This simple approach proposed here permits a highly integrated, low area, low power, and low-cost implementation.

3.1 Introduction

The DCR and low-IF architectures require linear LNA and a Mixer with a high frequency LO with precise quadrature outputs. In these types of receivers, the conventional approach of designing independently these blocks is not longer suitable. Alternatively, a co-design methodology for adapting the Mixer to the LNA and to the LO is required. All these requirements are difficult to fulfill simultaneously, and therefore, an optimized trade-off process should be followed.

In this work the Mixer plays a center role in a co-design strategy applied jointly to the LNA and LO for applications in the sub-gigahertz ISM-band and with low to moderate data rate, which can be applied to direct or low-IF receivers. The main objective is to avoid matching buffers in LNA and oscillator outputs, and directly connect them to the Mixer without using AC coupling capacitors and choke inductors.

3.2 Front End Building Blocks

The key blocks of the architectures presented in chapter 2 are the LNA, Mixer, Oscillator and ADC. A brief description of an LNA and LO to be used in the co-design strategy with the Mixer will be shown in this section. The following sections are dedicated to the Mixer architecture and dimensioning process. The ADC will be discussed in chapter 4.

3.2.1 LNA

The LNA, shown in figure 3.1, uses the source-degenerated topology around input transconductance transistor $M1$. This architecture is very common among narrowband LNA's as it is very close to achieving the goal of providing the input match and best noise performance simultaneously [2].

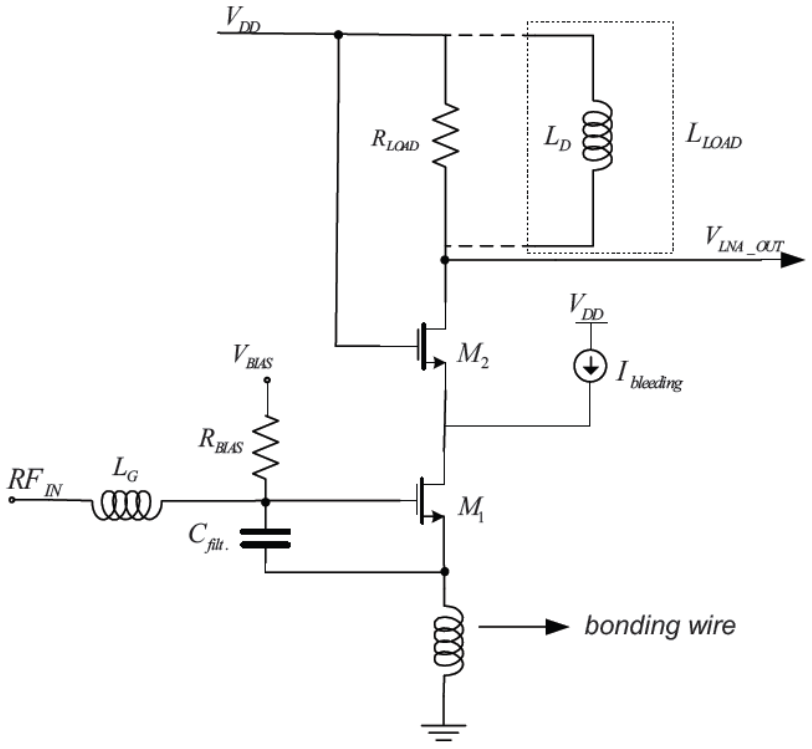


FIGURE 3.1: LNA circuit schematic [2].

Note that the input must be matched to 50 Ω due to the off-chip antenna.

From the traditional inductive load LNA (a $27nH$ inductor was used, one of the maximum available value from the chosen technology) with a 50 Ω output, a 15 dB gain can be

obtained with 2.7 dB of Noise Figure (NF) at 900 MHz. However, because the Mixer's inputs are MOS gates (meaning that it is driven by a voltage) it is the LNA voltage gain that should be considered. Since the output of the LNA does not require the 50 Ω match because all blocks are in the same chip, replacing this inductor by a 700 Ω resistor (which can not be higher due to power supply headroom) a 28 dB voltage gain with a 2.24 dB noise figure can be achieved. The power consumption is lower than 9 mW for the two cases [2].

3.2.2 Quadrature Local Oscillator

The schematic of the two-integrator oscillator [6], shown in figure 3.2, is realized by a differential pair (transistors M) and a capacitor (C). The oscillator frequency is controlled by I_{tune}.

There is an additional differential pair (transistors M_L), with the output cross-coupled to the inputs, which performs two related fuctions:

- compensation of the losses due to R to make the oscillation possible (a negative resistance is created in parallel with C),
- amplitude stabilization, due to the non-linearity (the current source I_{level} controls the amplitude).

To start the oscillations the condition $gm > 1/R$ must be met. Moreover, the I_{level} is used to control the output signals amplitude [2].

The oscillator frequency varies by changing either the capacitance or the transconductance. In a practical circuit one can use varactors to change the capacitance or, most commonly, change the tuning current (and therefore, the transconductance).

These oscillators have wide tuning range with very precise inherent differential quadrature outputs (less than one degree quadrature error), which are required for very compact DCR and low-IF receivers [6].

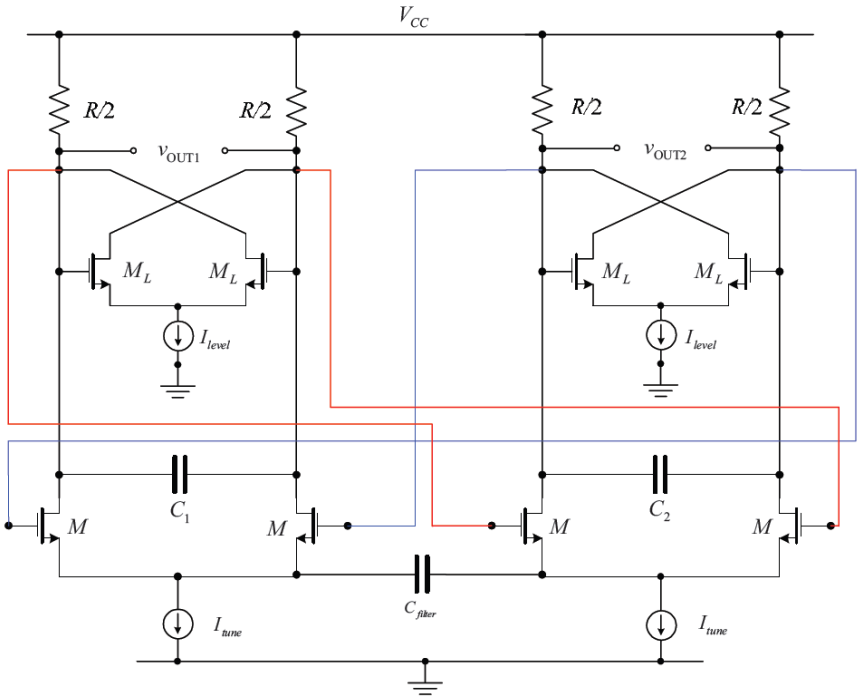


FIGURE 3.2: LO circuit schematic [2].

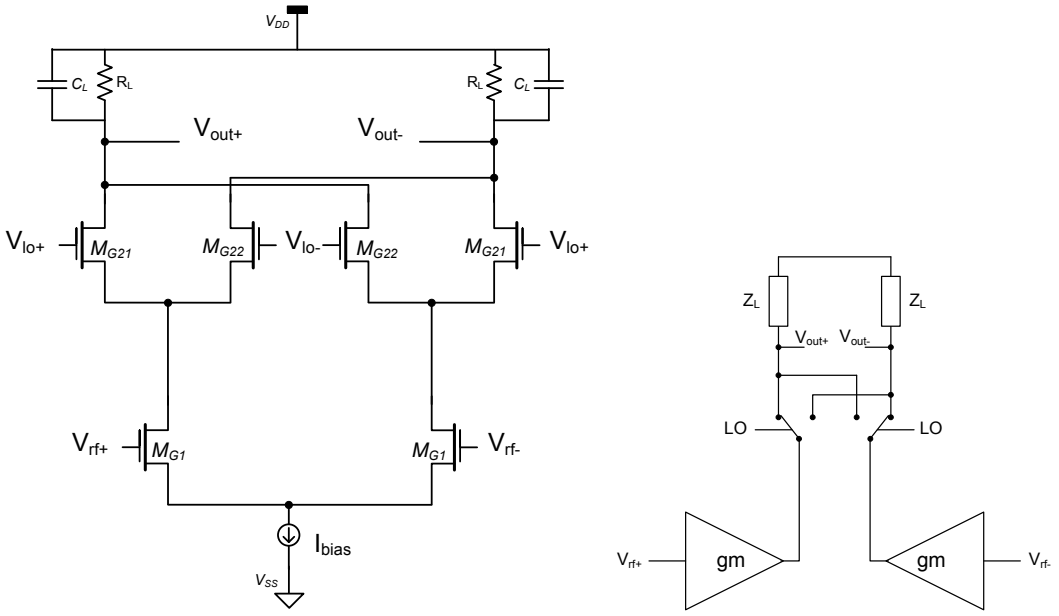
3.3 Mixer

The Mixer architecture is based in the double balanced Gilbert Cell presented in figure 3.3(a). It was initially used in bipolar technology as an analog multiplier.

This Mixer has higher gain, lower noise figure, good linearity, high port-to-port isolation, high spurious rejection, and less even-order distortion, with respect to the single-balanced. The main disadvantage is the increased area and power consumption. Additionally, it requires the use of a balun transformer to provide the RF differential at the Mixer input or a single-ended to differential LNA (a balun-LNA, this is an ongoing parallel work).

This circuit can be divided into four different sections. Apart from the bias current source, this sections can be related to the simplified model presented in figure 3.3(b)

- Bias - the I_{bias} current source defines the operating point for all the transistors;
- RF stage - transistors M_{G1} implement a differential transconductance gain stage in the v_{rf} signal;
- Switching stage - transistors M_{G21} and M_{G22} act as switches and are responsible for the multiplication;



(a) Circuit schematic

(b) Simplified model

FIGURE 3.3: Schematic and simplified model of the Gilbert Cell

- I/V conversion - the load resistor, R_L, provides a voltage output and C_L filters unwanted noise.

It works as follows. Assuming ideal switches in the switching stage, the output voltage v_{out+} is defined either by v_{rf+} or v_{rf-}, depending on the position of the switches (the same happens with v_{out-}). Since the switches are driven by the local oscillator, this switching in the v_{rf} polarity at the output produces the desired multiplication ($v_{out} = \pm gm\, v_{rf}\, R_L$, with the $\pm$ changing at the LO frequency).

To ensure that the M_{G21} and the M_{G22} switches are not on simultaneously, which is needed for proper operation, the LO waveform should be an ideal square wave. This is impossible to achieve and, in practice, there is a period of time when they are both on (when the LO voltage crosses close to the threshold voltage of the transistors). However, an high amplitude sine wave, which approximates to a square wave near the switching point, can be used [10] with no significant loss in performance. In [11] one can find an extensive work with different strategies for dimensioning the transistors and achieve key parameters for optimization. These were used as guide lines for the initial dimensions.

As mentioned, the purpose of this work is to apply a co-design strategy to the Mixer, LNA and LO so, no special effort was dedicated to noise figure or gain optimization.

3.3.1 Proposed Architecture

The adapted I/Q Mixer topology, shown in figure 3.4, is based in a active double balanced Gilbert cell. The obvious change to the traditional Gilbert's cell is the I/Q LO differential inputs. Note that the same RF stage is used for both the I and Q Mixers. This reduces area and minimizes the effects of mismatches.

The output DC level in the traditional Gilbert's Cell (which can limit the output swing) is defined by $V_{DCout} = V_{DD} - R_L \frac{I_{bias}}{2}$, and the voltage gain is proportional to $gm\, R_L$. To increase the gain one can increase either I_{bias} or $W_{M_{G1}}$ (which, in turns, increases gm) or increase the load resistor R_L. However, the output DC level is limited not only by the output swing, but also by the need of ensuring the saturation operating point in all the transistors.

Due to the low voltage supply, the bias current source has been removed to leave some more headroom for transistor biasing. This adds some difficulties because now there is no way to easily define the transistors operating point. However, the addition of two current sources, I_{bleed}, allow some control in this subject. Moreover, these current sources can be used to increase the Mixer's gain by increasing the gm of M_{G1} ($gm \approx \frac{2I_D}{V_{dsat}}$), without increasing the DC current on R_L, thus keeping the output DC unchanged.

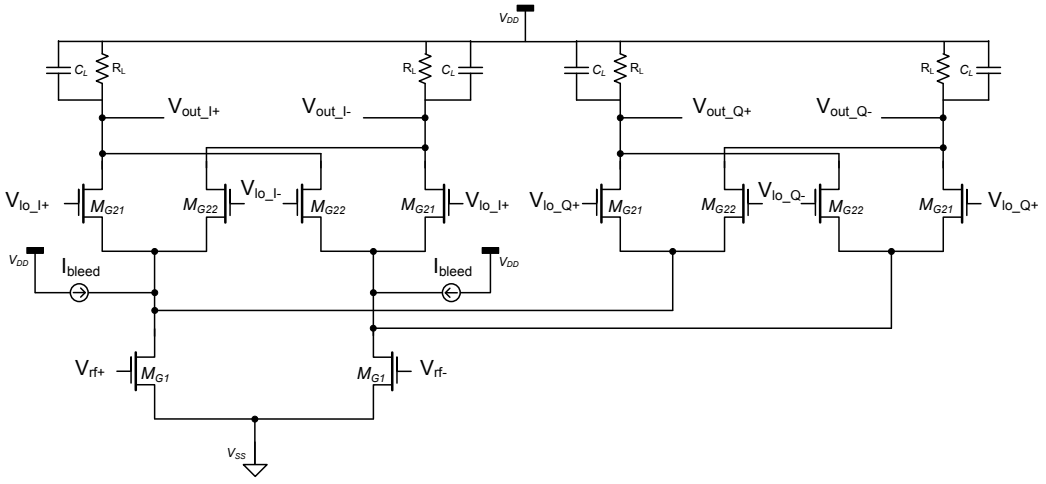


FIGURE 3.4: Proposed Mixer circuit schematic.

3.3.2 Design Methodology

The design process begins by maximizing the LNA voltage gain for a given input match criteria (50 Ω in this case). By its turn the oscillator is designed to maximize the signal output swing voltage and improve the I/Q signals accuracy for a given power. With the obtained DC components at the output of these blocks, the Mixer is then optimized to reach a reasonable conversion gain (CG) and noise figure. In order to reduce the total area, the design must remove as much as possible the use of inductors and AC coupling capacitors.

The following guide lines were used in the Mixer's dimensioning process [10, 11]:

- ensure that all transistors are in the saturation regime;
- dimensioning starts with equal width for all transistors;
- the noise figure is minimized by increasing the transistors width;
- increasing the width of M_{G1} has significantly more benefits in the noise figure;
- the Third-order Input Intercept Point (IIP3) is proportional to V_{dsat} of transistors M_{G1};
- flicker noise is minimized by smaller transistors in the switching stage;
- the conversion gain is proportional to $gm_{M_{G1}}R_L$;
- minimum length should be used for fast operation and gain.

The most important limitation is that, in a co-design strategy, one must consider not only the Mixer but also the LNA and LO. In this case, the main goal is to bias transistors M_{G1} with the DC level provided by the LNA, and transistors M_{G21} and M_{G22} with the DC level of the LO. This can be done because the Mixer's inputs are transistor gates (high impedance). This leaded to minor adjustments in the LNA and LO circuits, with acceptable performance losses when compared with the isolated optimization.

Due to the removal of a bias tail current, body effects in the switching stage and low voltage supply, the process of ensuring saturation in all transistors must be done by simulation. In an initial sizing, following the above guide lines, all the transistors were set to minimum length and equal width of $30\mu m$ and $R_L = 200\Omega$. After some iteration with simulation

results, the final parameters of the designed Mixer are: $R_L = 800\Omega$, $C_L = 2.5pF$, $(W/L) = 100\mu m/0.13\mu m$ for the switching stage transistors, $(W/L) = 30\mu m/0.13\mu m$ for the RF stage transistors and $I_{bleed} = 4.15mA$.

3.4 Simulation Results

To validate the proposed strategy, the LNA, Mixer and quadrature LO has been simulated using a 130 nm CMOS technology with $1.2\,V$ power supply. The following figures show the simulation results for the final circuit.

Remember that the LNA used is single ended, and due to the removal of the current source the Mixer architecture results in a pseudo-differential structure, not a fully differential. This means that the full potential of the double balanced architecture is not achieved and the following results are expected to be improved with a differential input.

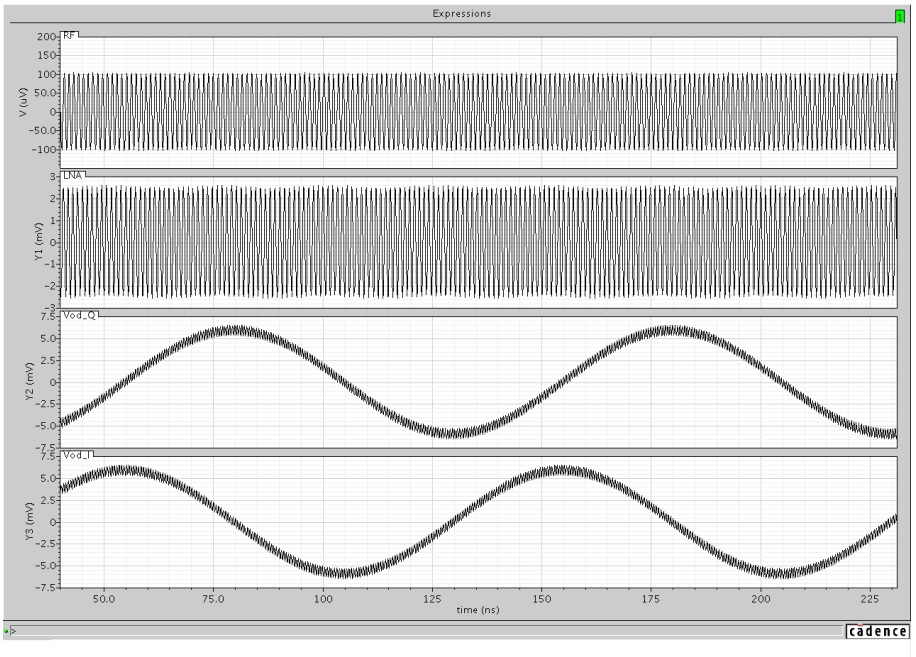


FIGURE 3.5: Simulated transient analysis.

Figure 3.5 shows the transient analysis with (from top to bottom) LNA input, LNA output/Mixer Input, Mixer's differential Q output and Mixer's differential I output.

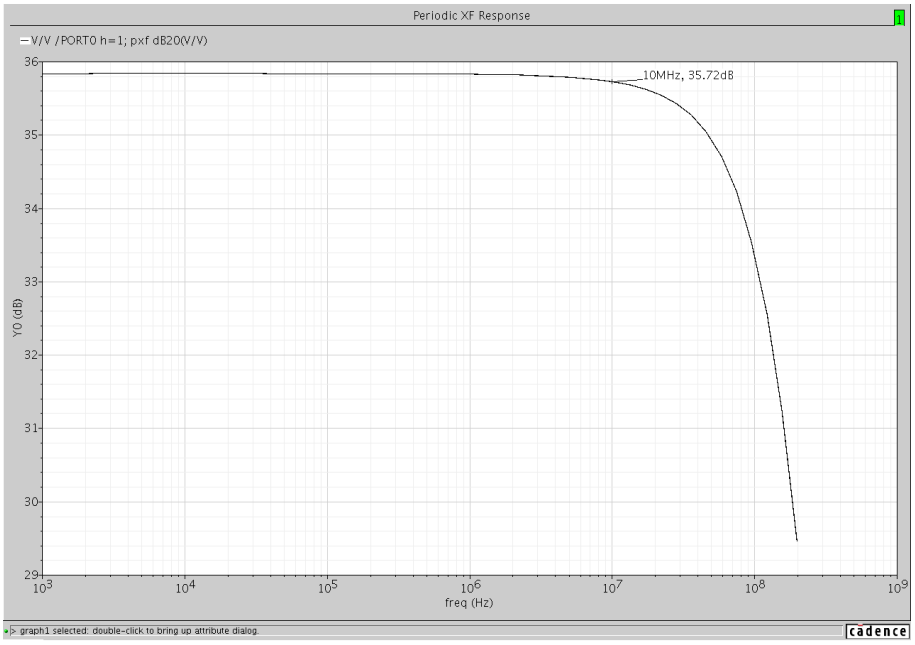


FIGURE 3.6: Simulated cascade Conversion Gain.

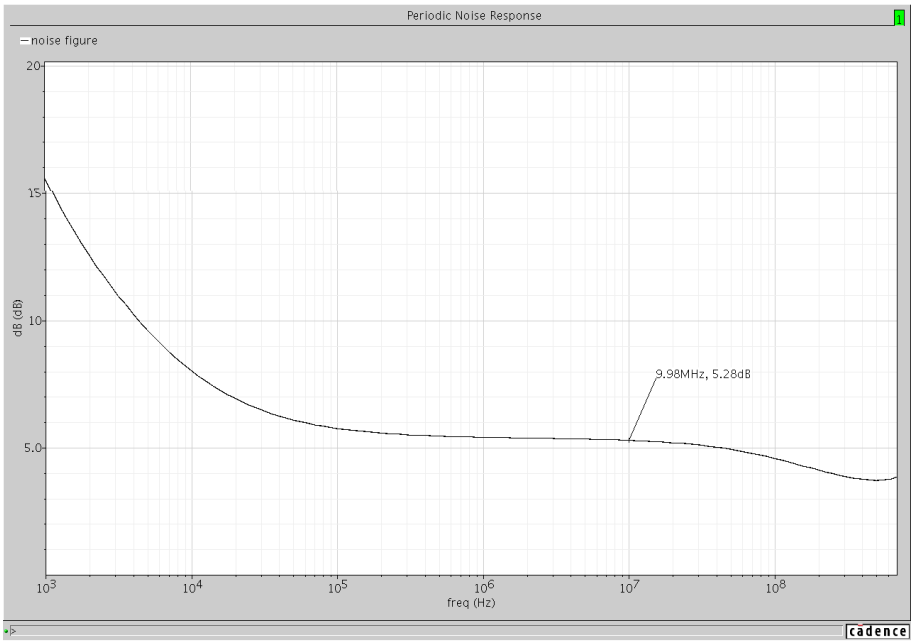


FIGURE 3.7: Simulated cascade Noise Figure.

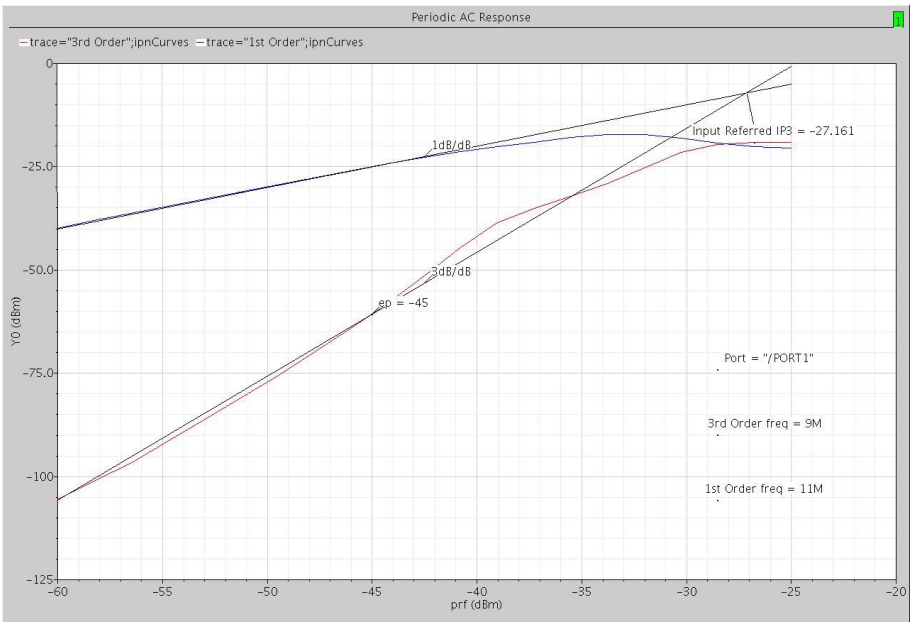


FIGURE 3.8: Simulated IIP3.

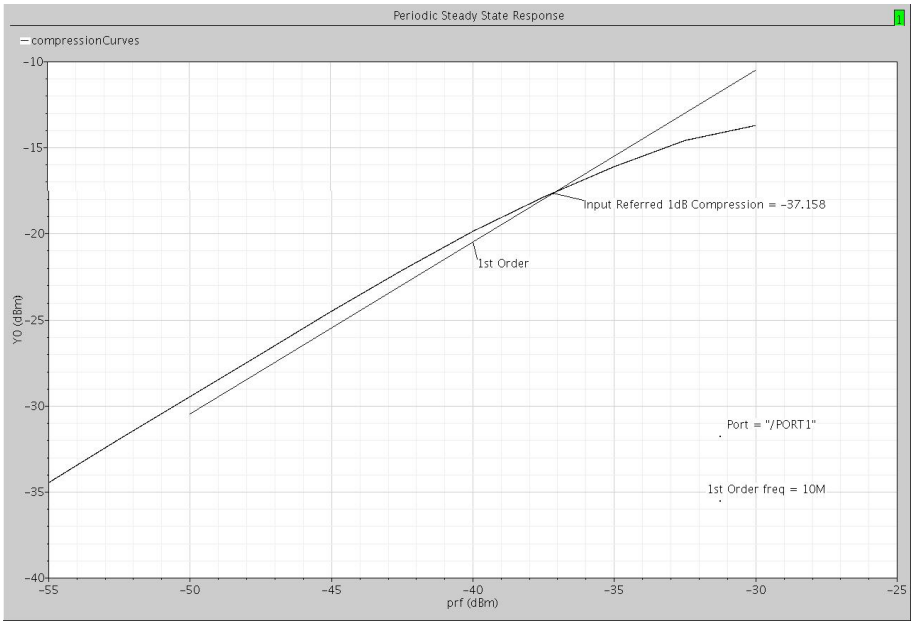


FIGURE 3.9: Simulated 1-dB Compression Point.

Two different simulation results are presented for the complete front-end, obtained from SpectreRF [12] simulator, using BSIM3V3 [13] models, including noise. In table 3.1, the LNA was optimized to achieve higher gain, restricted by co-design trade-offs. In order to improve linearity, the LNA gain was reduced by removing the bleeding current and adjusting the resistive load R_L to meet the same DC output (thus, keeping the Mixer and LO unchanged). Table 3.2 shows the simulation results obtained in this case.

Parameter @ $10MHz$	Value
NF	$5.28dB$
CG	$35.7dB$
$IIP3$	$-27.16dBm$
$1dB - CP$	$-37.16dBm$

TABLE 3.1: Simulated cascade results (Optimized for LNA gain).

Parameter @ $10MHz$	Value
NF	$6.16dB$
CG	$25.5dB$
$IIP3$	$-17.33dBm$
$1dB - CP$	$-28.95dBm$

TABLE 3.2: Simulated cascade results (Reduced LNA gain).

In both cases, the NF and CG results were obtained with the complete front-end models, but, due to simulator limitations, IIP3 and 1dB compression point (1dB-CP) were obtained using an ideal LO with the same amplitude and frequency simulated with the LO, while loading the Mixer. However, this should not translates in a significant change while measuring IIP3 and 1dB-CP.

3.5 Conclusions

In this work, a co-design strategy for the implementation of a low-voltage, low-area, low-cost, fully integrated CMOS receiver was presented. This approach avoids 50 Ω matching buffers and networks, AC capacitors coupling, and DC choke inductors.

The presented work uses a resistive load LNA, with 700Ω load and inductor-less differential RC quadrature oscillator, which are combined with a Mixer in a co-design strategy. A

current bleeding technique was applied at the LNA and Mixer, due to the low power supply voltage. The low area quadrature two-integrator oscillator uses a capacitive filtering technique, which reduces the oscillator phase-noise and the harmonic distortion.

The approach presented has only one inductor allowing the design of a very compact and low cost receiver (DCR or low-IF), which is required for low data rates ISM applications. The proposed receiver was designed and simulated in UMC 130nm CMOS technology. The total conversion voltage gain is $35.8\,dB$ and the cascade noise factor is $5.3\,dB$ for the interest band.

Chapter 4

Parametric Based Comparator for Flash ADCs

Analog to digital converters are used as an interface between the analog and the digital world, performing the transformation from continuous time and amplitude to discrete time and quantized amplitude.

In this chapter, the background theory necessary to understand ADCs in general will be dealt and such knowledge will be applied to the proposed architecture.

4.1 Introduction

The basic operation of an ADC can be summarized by the block diagram in figure 4.1.

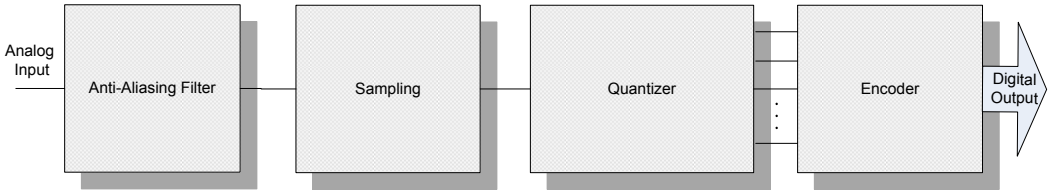


FIGURE 4.1: Block diagram of an ADC.

As mentioned above, the purpose of an ADC is to transform an analog signal to a digital one. With little detail, the analog signal, after *filtering*, is sampled at a given rate and maintained constant for a while by the *Sample&Hold* circuit (typically, it tracks the input signal during half the period and holds it for the rest of it). While constant, a *Quantizer*

circuit determines which discrete value is the most appropriate to represent the input signal. Finally, the *Encoder* selects an appropriate digital code.

There are several architectures for ADCs, such as Pipeline, Integrating, Successive Approximation, Flash, etc. The focus of this work will be on Flash converters, which are appropriate when a high-speed and low-resolution solution is the goal.

The main advantage of Flash converters is speed. This is obtained by a brute-force approach: the sampled input voltage is compared simultaneously (in parallel) by different comparators, each one set to compare the input voltage with a different quantization level. For a given input signal, the comparator output is "1" if the reference voltage is lower than the input voltage, and "0" otherwise, giving a thermometer effect at the comparators bank output. Figure 4.2 is a common representation of a generic Flash converter.

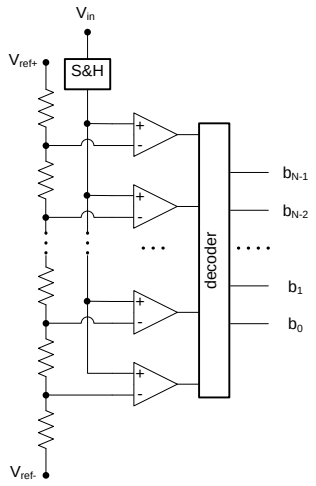


FIGURE 4.2: Block diagram of a Flash converter.

With this architecture the circuit gives one digital code at each clock period, which permits high-speed but, since it needs $2^n - 1$ comparators to achieve a n-bits comparator it has the disadvantage of power consumption and area. Also, due to metastability and offset in the comparators, a pre-amplification stage is needed before comparison, increasing complexity and power consumption if medium resolution is needed. This issues will be discussed in later sections.

4.2 Anti-aliasing Filter

Figure 4.1 shows that the first block in the signal path is an *Anti-Aliasing Filter*. This block will not be included in the proposed circuit. Nevertheless, it is important to, at least, discuss the importance of this block, since it will be needed in case of circuit implementation.

The sampling operation, in the ideal case, is a sequence of delta functions whose amplitude equals the input signal at the sampling times. This operation can be easily understood in figure 4.3.

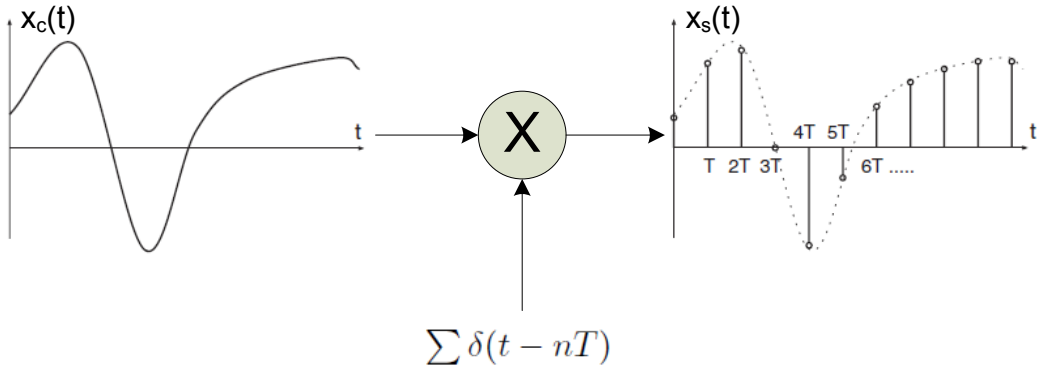


FIGURE 4.3: Ideal Sampler.

Mathematically, $x_s(t) = \sum x_c(t)\delta(t - nT)$, where $x_c(t)$ and $x_s(t)$ are the continuous and sampled signals, respectively, and $\delta(t)$ is the unit impulse function, also called *Dirac Delta function*. When $x_c(t)$ is sampled at $f_s = 1/T$, $x_s(t)$ takes the following form, in the frequency domain [14]:

$$X_s(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X_c(j2\pi f - jk2\pi f_s) \tag{4.1}$$

Equation 4.1 shows that the spectrum of a sampled signal, $x_s(t)$, equals a sum of shifted spectra of $x_c(t)$. These replicas are centered at multiples of the sampling frequency f_s, being shifted by kf_s, resulting in a periodic spectrum with infinite replicas of the original signal.

The first conclusion that results from equation 4.1 is that the sampling frequency has to be chosen in such way that the different replicas do not overlap. This is only possible by

assuming that the input signal is band limited. Figure 4.4 c) shows an example of this overlap effect.

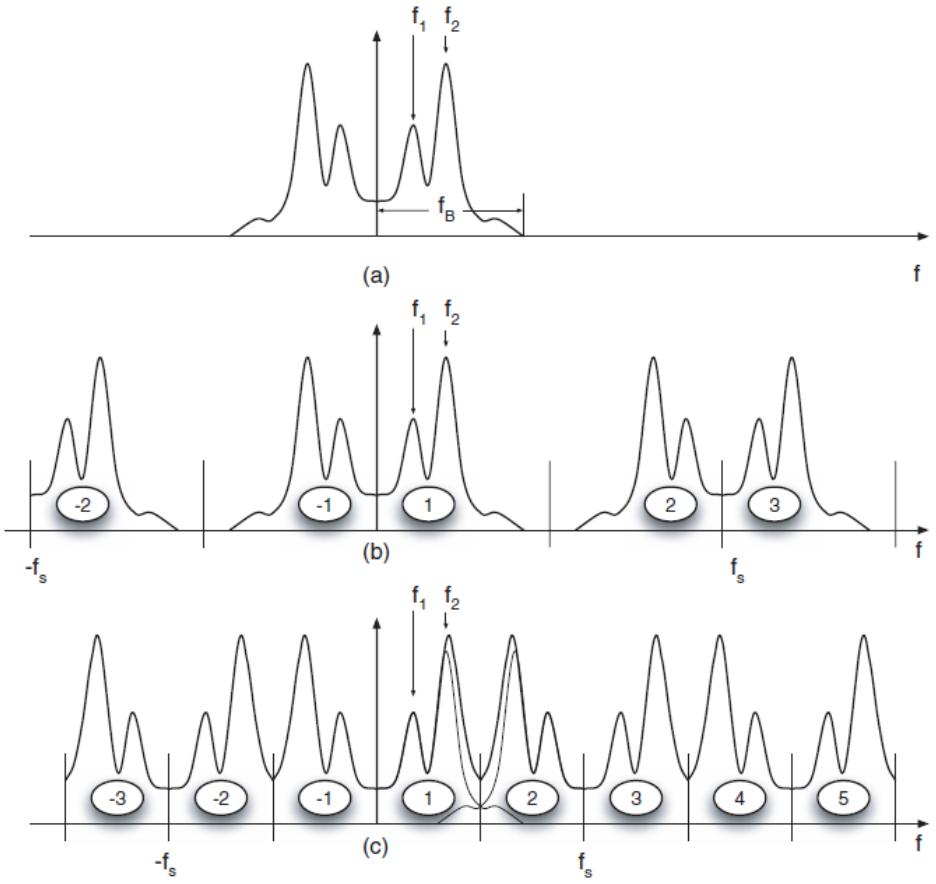


FIGURE 4.4: a) Bilateral spectrum of a continuous-time signal. b) Sampled spectrum by using $\frac{fs}{2} > f_B$. c) Sampled spectrum with $\frac{fs}{2} < f_B$ [15].

Even if the signal is band limited, noise and unwanted signals will be present, scattered across the frequency spectrum, and if not filtered, their spectrum will fold into the band of interest and corrupt the signal band. This effect is known as aliasing and can be avoided if the Nyquist frequency[1] (or $\frac{f_s}{2}$) is greater than the bandwidth, or maximum component frequency, of the signal being sampled, as 4.4 b) illustrates.

The second, and not so obvious, conclusion is that an higher sampling frequency allows relaxation on filter specifications. The filter must have an effective stop-band starting at $f_s - f_B$ in order to avoid overlapping, so the width of the filters transition-band is from f_B

[1]The Nyquist frequency should not be confused with the Nyquist rate, which is the lower bound of the sampling frequency that satisfies the Nyquist sampling criterion for a given signal or family of signals, ie, a signal can be fully described by an uniform sampling at $f_{Nyquist} = 2f_B$.

to $f_s - f_B$. Higher sampling frequency implies wider transition-band and, consequently, a lower order filter.

The design of the anti-aliasing filter is critical. Poor filtering will affect the overall performance and can't be neglected in the implementation.

4.3 Sample & Hold With Embedded Parametric Pre-Amplification

A simple representation of a Switched Capacitor (SC) that preforms the Sample&Hold (or Track&Hold[2]) operation is presented in figure 4.5.

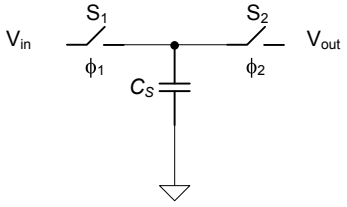


FIGURE 4.5: Basic Sample&Hold.

The Sample&Hold operation is realized in two phases. In the first phase (sample), the input switch S_1 is *on* and the output switch S_2 is *off*. During this phase the capacitor C_S is charged to the input voltage. In the second phase (hold), S_1 goes *off* and S_2 switches *on*. This leaves the top plate of the capacitor floating, retaining the charge (proportional to the input voltage, $Q = CV$). Notice that sampling occurs at the end of the sample phase, when the input switch goes *off* and freezes the capacitor charge. In addition, an output buffer can be used to prevent unwanted discharge of the capacitor during the hold phase and to provide a voltage output.

It is important to say that some attention must be taken in order to prevent switches S_1 and S_2 to be *on* simultaneously. In CMOS implementations, this is usually solved by using non-overlapping clock phases and/or very sharp clock-edges to drive the switches.

Also, by noticing that a non-ideal input switch has finite conductance when is *on*, the circuit is a RC network. It is important to keep in mind that the time constant ($\tau = RC$) of this network must be lower than the time allowed to charge the capacitor (how much lower depends on the required accuracy).

[2]When the output is available during the sample phase a Sample&Hold is named Track&Hold.

4.3.1 Switched Capacitor Sampling Circuit

To analyze a switched capacitor circuit presented in figure 4.5 one should remember the *Charge Conservation Principle* which states that electric charge can neither be created nor destroyed. Thus, in the ideal case, when the circuit switches from the sample phase to the hold phase, the charge stored in C_S remains constant.

At the end of the sample phase, the charge stored in C_S is given by:

$$q_{in} = C_S \, V_{in}$$

and, at the beginning of the hold phase:

$$q_{out} = C_S \, V_{out}$$

Considering that there is no path for charges to flow out from C_S, q_{out} remains constant during the hold phase.

Thus, the charge transfered from the input to the output in each cycle is given by:

$$\Delta Q = q_{out} - q_{in} = C_S \, (V_{out} - V_{in}) \tag{4.2}$$

Notice that, when applying the charge conservation principle, $\Delta Q = 0$ and equation 4.2 results in $V_{out} = V_{in}$, as wanted at this point.

Moreover, if V_{in} and V_{out} are DC voltage sources, since the charge transfer is repeated every clock period, one can find the equivalent average current by multiplying equation 4.2 by the clock frequency f_s [14]:

$$I_{avg} = \Delta Q \, f_s = C_S \, (V_{out} - V_{in}) \, f_s \tag{4.3}$$

This result leads to the realization that the switched capacitor can be modeled as a resistor R_{eq}:

$$R_{eq} = \frac{V_{out} - V_{in}}{I_{avg}} = \frac{1}{C_S f_s} \tag{4.4}$$

The equivalent resistor model for the SC is valid for as long as the I_{avg} concept is, i.e., for f_s much higher than the input signal frequency the input sources can be approximated by DC sources and the result stands. If this is not the case, Discrete-Time analysis is required. Nevertheless, this result helps to understand how a SC network can be used to replace the resistor ladder in the Flash architecture (figure 4.2). This will be dealt in section 4.5

4.3.2 Switches

The MOS transistor is a natural switch. When the gate-source voltage, V_{gs}, is below the threshold voltage of the transistor, V_{th}, there is no conductive channel connecting the source and drain, i.e., the transistor is an open switch. If V_{gs} exceeds V_{th}, a conductive channel is created and the transistor implements a closed switch. Neglecting body effect, which will be briefly discussed at the end of this section, and assuming that the drain-source voltage is low or zero, the transistor operates in the triode region and, for an NMOS transistor:

$$I_D = \mu_n C_{ox} \left(\frac{W}{L}\right) \left[\left(V_{gs} - V_{tn}\right) V_{ds} - \frac{V_{ds}^2}{2}\right] \tag{4.5}$$

Neglecting second order effects (remember that V_{ds} is assumed small so $V_{ds}^2 \approx 0$) the *on*-conductance is given by:

$$G_{on} = \frac{I_D}{V_{ds}} = \mu_n C_{ox} \left(\frac{W}{L}\right) \left(V_{gs} - V_{tn}\right) \tag{4.6}$$

Equation 4.6 shows that, for a fixed L[3], the *on*-conductance increases with the transistor's width. This means that increasing W will decrease τ of the RC network.

Equation 4.6 also shows that the *on*-conductance goes to zero when $V_{gs} - V_{tn}$ goes to zero. Notice that, in a switch, $V_g \approx V_{DD}$ (or V_{SS} in a PMOS), which is given by the clock signal, and $V_s = V_{in}$[4]. This means that when the input voltage is contained within

[3]Minimum L (L_{min}) will always be used to achieve faster transistors, unless the short-length effects are considerably relevant.

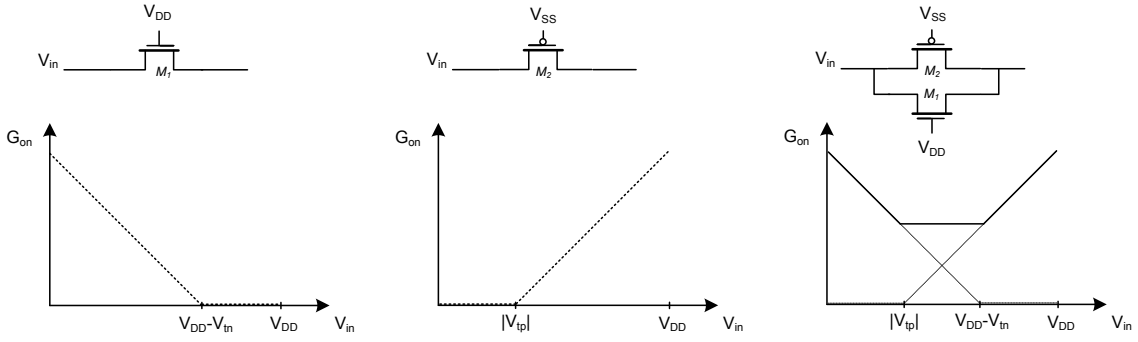
[4]Since there is no physical distinction between drain and source in a MOS transistor, in a NMOS the terminal at lowest potential is the source. In the PMOS, the source is the terminal at higher potential. This distinction is not relevant in this case because we are assuming that V_{ds} is very small

suitable values, one can use a single transistor (NMOS or PMOS depending on the input voltage). In particular, NMOS transistors fail when $V_{in} > V_{DD} - V_{tn}$ and PMOS fail when $V_{in} < |V_{tp}|$.

When the input voltage varies over a large range, a transmission gate, formed by a pair of transistors (NMOS and PMOS), ensures that when one has zero *on*-conductance due to low V_{gs}, the other is working properly. The resulting *on*-conductance in this case is given by $G_{on} = G_{on_p} + G_{on_n}$, and is less dependent of the input voltage.

Moreover, since $\mu_n \approx 4\mu_p$ (the exact value is technology-dependent), PMOS transistors must be approximately four times wider then NMOS in order to achieve the same *on*-conductance.

Figure 4.6 is a generic representation of the *on*-conductance in CMOS switches as a function of V_{in} (neglecting body effect and assuming that the clock generator can provide either V_{DD} of V_{SS} to the transistor's gates).



(a) G_{on} in an NMOS (b) G_{on} in an PMOS (c) G_{on} in a Transmission Gate

FIGURE 4.6: *On*-conductance in CMOS switches as a function of V_{in}

Note that in the PMOS switch, the transistor's width must be approximately four times the NMOS width in order to achieve the symmetry illustrated in figures 4.6(a) and 4.6(b).

Also, note the importance of supply and threshold voltages, particularly in the transmission gate in figure 4.6(c). Ideally, the input voltage should be between $|V_{tp}|$ and $V_{DD} - V_{tn}$ in order to keep G_{on} constant (the flat horizontal line). For such input voltage values (or input swing), the *on*-conductance is input-independent[5]. A larger input swing can

[5]Remember that body effect is being neglected so far.

be achieved by increasing the gate's voltage or by lowering the threshold voltage. Unfortunately, this two are tied together, i.e., as technology evolves, both the supply and threshold voltages decreases, but not at the same rate. In recent technologies, the linear behavior of switches is restricted to a very small input swing.

Charge pumps, switch bootstrapping and double supply voltages [15], are other methods used to overcame the above mentioned problems. The approach here is to increase the gate voltage to values higher than V_{DD} when the switch is *on*. This is done by charging a capacitor during the off period in the former two cases, or by using a second voltage source in the circuit in the latter. Added complexity, reduced life-time due to higher V_{gs} and costs are strong arguments to avoid this solutions if a simple pair of transistors can do the job.

The finite *on*-conductance of CMOS switches must be taken into account in the determination of the settling time, which will define the Sample&Hold's accuracy.

Unfortunately, CMOS switches also suffer from other phenomenas. One particularly important in Sample&Hold circuits is the *Clock Feed-Through* and *Charge Injection*, explained next.

Clock Feed-Through and Charge Injection

As already discussed, if V_{gs} exceeds V_{th}, a conductive channel, formed by accumulated minority carriers under the gate, is created[6], and a transistor implements a closed switch. When the transistor switches off, the channel's charges must be removed from under the gate by flowing through the source and drain terminals. This results in charge injected into the sampling capacitor that is a fraction of the total channel's charge.

The channel charge of a transistor in the triode region is given by:

$$Q_{ch} = WL\, C'_{ox}\, (V_{gs} - V_{tn}) \tag{4.7}$$

The fraction of Q_{ch} injected into the sampling capacitance depends on the MOS parameters, the clock transition time and boundary conditions on both sides (drain and source) of the transistor [15]. A common approach is to assume sharp clock transitions and equal

[6] In this conditions, the channel is said to be inverted.

boundary conditions. In this case the channel's charge flow evenly through the source and drain and the charge injected into the sampling capacitance is $\frac{Q_{ch}}{2}$.

To minimize this type of injected charges by a switch, an additional *Dummy* switch is used, as illustrated in figure 4.7.

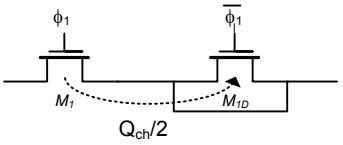


FIGURE 4.7: NMOS Switch and Dummy Switch to minimize Clock Feed-Through and change injection.

In this configuration, when M_1 switches off, M_{1D} turns on. The injected charges from M_1 are used in M_{1D} to create it's conductive channel. In the ideal case, all the injected charges are consumed by the dummy transistor, so the Dummy switch's width must be a fraction on the input switch corresponding to the fraction of the M_1 charges that would flow for the sampling capacitance.

For example, if the simplification of sharp clock edges and equal boundary conditions assumed before could be made, then one should have $W_{M_1} = 2\, W_{M_1 D}$. When using transmission gates, two dummy switches (one NMOS and one PMOS) should be used.

Moreover, overlap between the gate and drain terminals also results in parasitic coupling, thus injecting more charges in the sampling capacitor when the gate voltage changes. This effect is less severe then the one formerly described [14].

The Bulk-switching Technique

In the transistors shown in figure 4.7 the *bulk* (or substrate) connection is not represented. This is an usual graphic simplification which assumes that the transistors's bulk is connected to the lowest potential in the circuit (V_{SS}) in the NMOS case. For the PMOS transistors, the bulk is usually connected to the highest potential (V_{DD}). The physical reason for this fourth terminal is beyond the scope of this work.

However, it is important to refer that an important effect arises if the bulk's terminal voltage is different then the source voltage. This is named *Body Effect* and is a second order effect modeled as an increase in the threshold voltage, V_{th}, as the source-to-substrate

reverse bias voltage, V_{sb}, increases [14]. Equation 4.8 shows how the threshold voltage is related to the source-to-substrate voltage in an NMOS transistor.

$$V_{tn} = V_{tn0} + \gamma \left(\sqrt{V_{sb} + |2\phi_F|} - \sqrt{|2\phi_F|} \right) \tag{4.8}$$

where V_{tn0} is the threshold voltage with $V_{sb} = 0$. ϕ_F and γ are technology dependent constants.

In a switch, one can clearly realize that this effect is present because the source voltage is the input voltage. As a result, equation 4.6 takes the following form:

$$G_{on} = \mu_n C_{ox} \left(\frac{W}{L} \right) \left[V_{gs} - V_{tn0} \underbrace{-\gamma \left(\sqrt{V_{sb} + |2\phi_F|} - \sqrt{|2\phi_F|} \right)}_{decreased\ conductance} \right] \tag{4.9}$$

In typical CMOS technology, the entire chip's substrate is a p^- type[7], meaning that it is suited only for NMOS transistors. When one includes a PMOS transistor, which requires a n^- substrate, a well has to be made in the p^- substrate and filled with n^- material.

Since the NMOS transistor's substrate is common to the entire chip, it has to be connected to the lowest voltage potential, V_{ss}. This leads to body effect in all NMOS switches, which degrades conductance in a signal dependent form, as depicted in equation 4.9. However, in PMOS transistors, the substrate connection is individual. This means that different PMOS' substrate terminals in the circuit can be at different potentials.

This degree of freedom in PMOS transistors allows the use of the *Bulk-Switching* technique, as shown in figure 4.8. When the switch is *on*, transistor M_{1SB} connects the source and bulk terminals ($V_{SB} = 0$), thus eliminating body effect[8]. In the *off* position, M_{2SB} connects the bulk to V_{DD} to prevent positive source-to-bulk voltage [16].

[7] Physical layout of transistors is not a subject for this work. One can find such information in [14].
[8] In fact, the body effect is not entirely eliminated due to the fact that M_{1SB} is not an ideal switch.

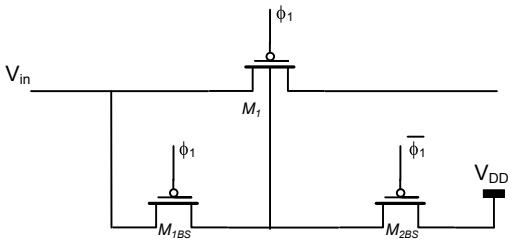
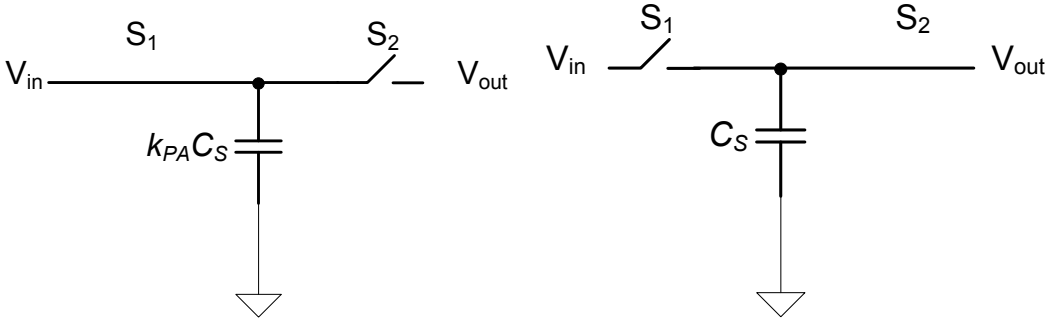


FIGURE 4.8: PMOS Switch with Bulk-Switching

4.3.3 MOSFET-based Parametric Amplification

A parametric amplifier is a circuit in which amplification is achieved by the use of variable (time-dependent) parameters or circuit elements [17]. In the MOSFET-based parametric amplifier the amplification is achieved by changing the sample capacitance parameter in an Switched Capacitor circuit. The principle can be easily understand by analyzing the circuit in figure 4.9.



(a) Switched capacitor in the sample phase

(b) Switched capacitor in the hold phase

FIGURE 4.9: Parametric amplification principle

As mentioned in section 4.3.1, and referring to figure 4.9, the charge stored at the end of the sample phase is given by $q_{in} = k_{PA}\, C_S\, V_{in}$, while, at the beginning of the hold phase is $q_{out} = C_S\, V_{out}$.

Thus, the charge transfered from the input to the output in each cycle is given by:

$$\Delta Q = q_{out} - q_{in} = C_S\,(V_{out} - k_{PA}\, V_{in}) \tag{4.10}$$

Applying the charge conservation principle in this case leads to $V_{out} = k_{PA}\, V_{in}$.

Figure 4.10 shows a four-terminal NMOS transistor with the source and drain connected to implement a three-terminal varactor[9]. Changing the control terminal (source and drain) from ground to V_{PULL} changes the operating point of the transistor, thus changing the total gate capacitance by a factor of k_{PA}.

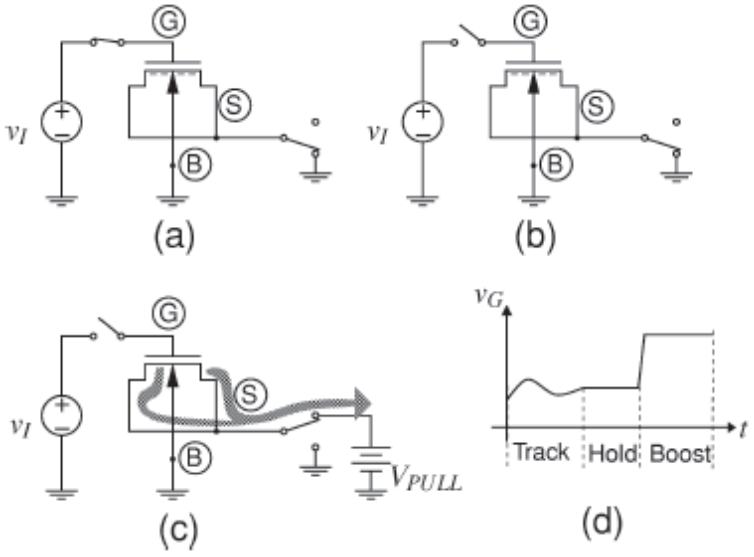


FIGURE 4.10: MOSFET parametric amplifier: a)Sample (Track) phase, b)Hold phase, c)Boost phase. d) Shows the gate voltage. [17]

In the sample phase (figure 4.10 (a)), the source and drain terminals are grounded and the input voltage is tracked in the gate. Positive charges accumulate at the gate and, to achieve charge neutrality, negative charges accumulate at the substrate. This charges are separated by the oxide layer (which forms the capacitor dielectric), as illustrated in figure 4.11(a).

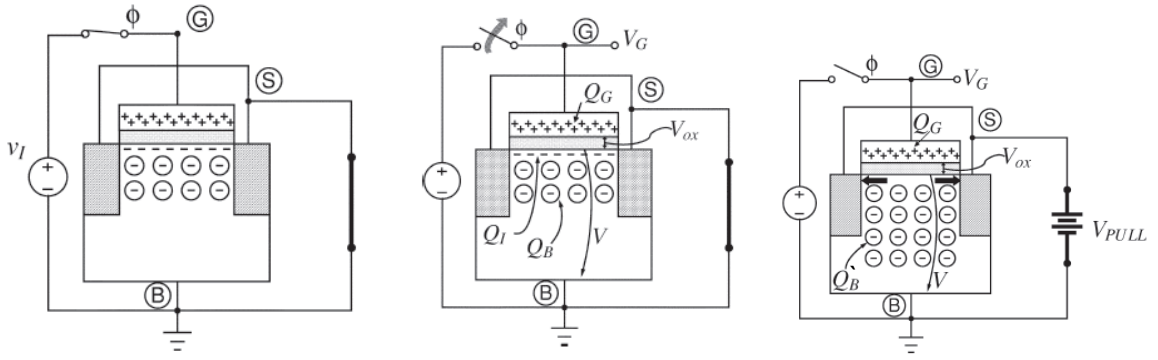
Assuming that V_G is large enough to drive the transistor into strong inversion, a fraction of this negative charges, Q_I, is used to create the inverted channel (represented by the minus signals right under the gate in figures 4.10 and 4.11 in sample and hold phases) and the rest are for the depletion region, Q_B. Thus, in this configuration:

$$Q_G = |Q_I| + |Q_B| \tag{4.11}$$

The voltage V_G is given by:

$$V_G = V_{ox} + V \tag{4.12}$$

[9]An adjustable reactance is usually referred as varactor.



(a) Sample phase (b) Hold phase (c) Boost phase

FIGURE 4.11: Cross section of the NMOS transistor and charges associated with the different phases [17]

where

$$V_{ox} = \frac{Q_G}{C_{ox}} \tag{4.13}$$

is the voltage drop across the oxide, $V = f(Q_B)$ is the voltage drop from the oxide-semiconductor interface to the bulk terminal and C_{ox} is the total gate capacitance ($C_{ox} = C'_{ox} \times WL$, with C'_{ox} representing the oxide capacitance *per* unit of area). V_G is held when the input switch turns off in the hold phase, as shown in figure 4.10 (d).

In the boost phase (figures 4.10 (c) and 4.11(c)), the large voltage V_{pull} is connected to the control terminal and the charges that form the inverted channel, Q_I, are pulled from under the gate. For this configuration:

$$Q_G = |Q'_B| \tag{4.14}$$

Note that, since the gate is now floating, Q_G remains constant and Q'_B must be higher than Q_B to compensate for the removed inversion charges and preserve charge neutrality. As a result, $V = f(Q_B)$ increases.

Moreover, equation 4.13 shows that the voltage drop across the oxide, which is only dependent of the gate charges, remains constant. Finally, from equation 4.12, one can realize that V_G increases.

A more intuitive way to look at this process is to realize that the inversion charges create a path connecting the grounded source and drain terminal to the lower oxide semi-conductor interface during the sample phase. Thus, the two "plates" of the sampling capacitor are separated only by the oxide layer and the capacitance of interest is C_{ox}. By removing the inversion charges, the two "plates" become separated by the oxide and the substrate layers, and the resulting capacitance is C_{gb}, which is smaller than C_{ox}. The resulting gain is

$$k_{PA} = \frac{C_{ox}}{C_{gb}}. \tag{4.15}$$

Equation 4.15 refers to an unloaded gain. In practical circuits this is not the case because, at least, parasitic capacitances exists loading the gate terminal. For this situation, the obtained gain becomes

$$k_{PA} = \frac{C_{ox} + C}{C_{gb} + C}. \tag{4.16}$$

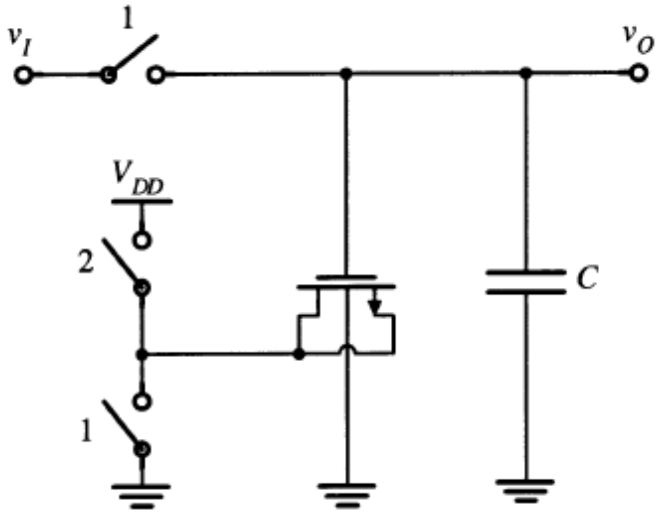


FIGURE 4.12: MOS capacitor amplifier [18]

A more detailed work in this subject can be found in [18]. Here, the analysis of the circuit in figure 4.12 leads to:

$$v_O = k_0 v_I + k_1 - k_2\sqrt{k_3 v_I + k_4} \tag{4.17}$$

Where $k_0..k_4$ are functions of C and technology parameters and can be found in [18]. For the purpose of this work, k_0 is the most relevant and is given by:

$$k_0 = 1 + \frac{C_{ox}}{C} \tag{4.18}$$

which can be related to equation 4.16 if one considers $C >> C_{gb}$.

From equation 4.17 one can realize that the main contributor for the parametric amplification gain is k_0. Moreover, this is higher than 1. Since C'_{ox} is a technology-dependent parameter, the gain of the circuit is defined by C, which can be a load capacitance but can simple be parasitic capacitances, and by the physical dimensions of the MOS device.

The charge transferred from the transistor gate to the capacitor C during the boost phase leads to larger voltage variations when it is small, as equation 4.18 indicates. But the jump in the common-mode component of the output voltage increases for smaller C. This is, therefore, a limitation to the maximum gain. The simultaneous use of an NMOS and a PMOS transistor alleviates this problem because the common-mode charges leaving the gates of these two transistors are of opposite signs and cancel each other [18].

The motivation behind parametric amplification is that, in principle, does not add any noise into the sample data. Moreover, the power dissipated is only dynamic and is the power required to pull out the inversion charges from under the gate during the boost phase [17].

4.3.4 kT/C Noise

Despite the parametric amplifications does not add, *per si*, any noise, there are additional noise sources evolved in the amplifier. One of the major contributers is usually referred as KT/C noise (see appendix B.4).

To improve the ADC's SNR, the sampling capacitance can be increased but, in integrated circuits, large capacitors may require a large die area, thus increasing costs. This is another motivation for parametric amplification because it is possible to achieve higher capacitance values with much smaller transistors than with the traditional MIMCaps (Metal-Insulator-Metal Capacitors) provided by the technology.

4.4 Dynamic Latch as a Comparator

The Dynamic Latch shown in figure 4.13 is a fast CMOS comparator. However, circuit-parameter deviations and charge injection mismatches between switches S_1 and S_2 causes a large input offset, limiting its resolution to about 5 bits [19]. To overcome this limitation, two approaches are used: reduce offset by circuit "improvement", and linear amplification preceding the dynamic latch so that the voltage applied to the latch is large enough to overcome this offset.

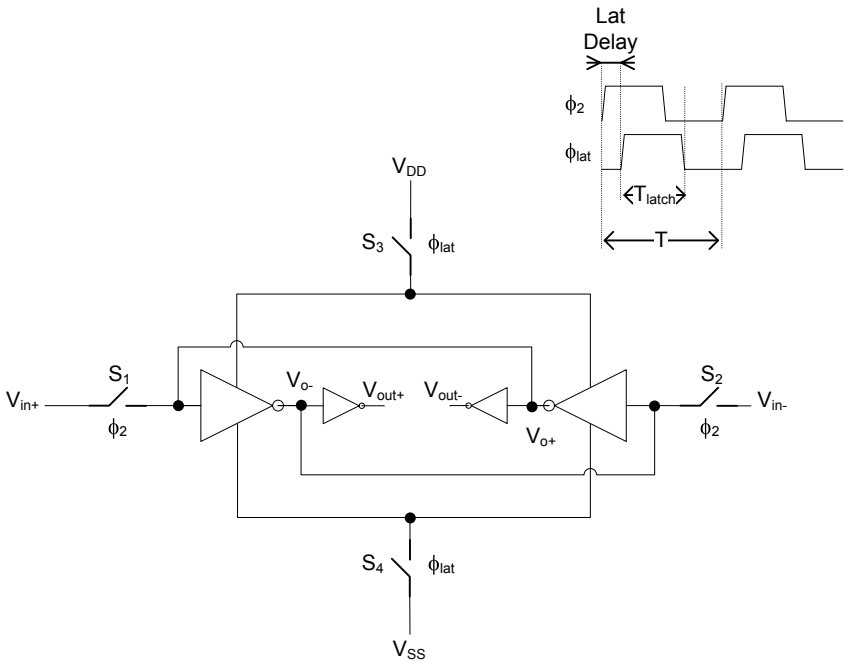


FIGURE 4.13: Dynamic Latch circuit.

Also shown in figure 4.13 is the clocking scheme, where there is no ϕ_1 as one should expect. This is to avoid confusion and allow to keep the same phase name throughout the entire text. It is slightly different that the one normally used due to the addiction of the ϕ_{lat} phase, which is a delayed version of ϕ_2.

In the traditional way, ϕ_{lat} is replaced by $\overline{\phi_2}$. When ϕ_2 is high, S_1 and S_2 are *on* and S_3 and S_4 are *off*. This phase can be called *Sensing Phase* as the latch senses the inputs V_{in+} and V_{in-}. Then, when ϕ_2 goes low, S_1 and S_2 switch off and S_3 and S_4 turns on. This isolates the input nodes from the latch and initiates positive feedback regeneration in the inverters, leading the outputs V_{o+} and V_{o-} to opposite rail voltages. At the end of this *Latch Phase*, the comparator's output are valid. Each output is followed by a buffer (inverter) to isolate the latch outputs from the following logic.

This approach minimizes the effects of Kickback, which denotes for the charge transfer either into or out of the input sources when the positive feedback is enabled, and the charge injection when S_1 and S_2 switches off [14]. This effect translates into a noise source and appears as strong glitches at the input voltage sources. Since the input nodes are isolated when the *Latch phase* starts, the kickback effect is strongly reduced. Nevertheless, the injected charge remains a problem.

To understand the clock scheme presented in figure 4.13, one must know that the Dynamic Latch's is preceded by a Sample & Hold With Parametric Amplification circuit (described in section 4.3, which samples during the "missing" ϕ_1 phase). The complete comparator architecture will be discussed in section 4.5.

So, the *Sensing Phase* of the Dynamic Latch corresponds to *Sample Phase* of the Sample & Hold and the beginning (the *Lat Delay* amount of time) of the *Hold Phase*. When switches S_3 and S_4, driven by ϕ_{lat}, turn *on*, the *Latch Phase* begins.

The reason for the *Lat Delay* time is to allow charge re-distribution at the input nodes, as well as to give time for the inversion charges to be pulled out from under the gate when parametric amplification is applied (see section 4.3.3).

Note that the Kickback effect in this case is reduced not by turning *off* the input switches in the latch, but by the input switches in the Sample & Hold circuit. Again, the injected charges problem remains as unwanted charges flowing into or out of the Sample & Hold capacitor.

4.4.1 Hysteresis

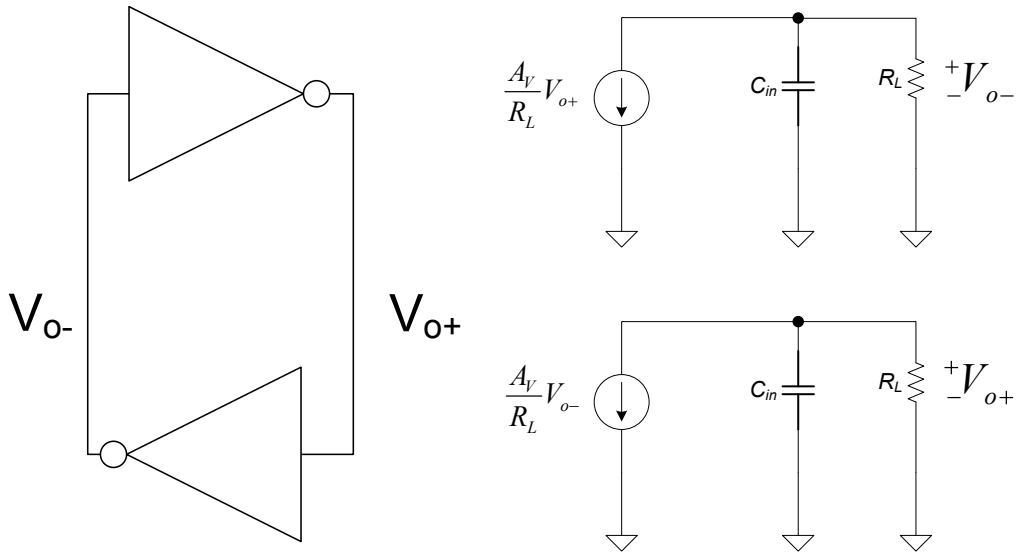
One very important issue in comparators is the tendency to "memorize" the output between decision cycles. If the comparator's outputs toggle in one direction in one cycle, the next cycle is affected by charges accumulated at the output nodes. This effect is sometimes called *Hysteresis* and results in different comparison levels that depend the previous state.

As an example, consider that the input differential voltage is negative and the comparator is set to toggle around $0\,V$. This conditions lead for an output state "0". If the input differential voltage increases, the comparator toggles to "1" at, lets say, $1\,mV$ (due to mismatches, for example). But if, from this output state "1", the input voltage decreases, it may toggle back to "0" only at $-2\,mV$.

To minimize this effect, one can reset the latch by connecting internal nodes (V_{o+} and V_{o-}) to one of the power supplies. This ensures that the starting state is always the same, thus eliminating memory. Moreover, by reseting the internal nodes, the comparator is set for its *trip point*, which speeds up the process when the comparator resolves small input signals [14].

4.4.2 Time Constant

According to [14], the time constant of the dynamic latch while in the *Latch Phase* can be found by analyzing the simplified circuit consisting in two back-to-back inverters, as shown in figure 4.14(a). If the internal nodes are reseted after each decision, as mentioned in section 4.4.1, one can consider that the outputs of the inverters are close to each other at the beginning of the latch phase. In this situation, the linearized model presented in figure 4.14(b) can be used.



(a) Simplified model of a Dynamic Latch in the latch phase

(b) Linearized model of a Dynamic Latch in the latch phase

FIGURE 4.14: Dynamic Latch's models in latch phase

Following [14], the minimum necessary time for the comparator to properly decide the correct output is given for:

$$T_{latchmin} = \frac{C_L}{G_m} ln\left(\frac{\Delta V_{logic}}{\Delta V_o}\right) = K\frac{L^2}{\mu_n\left(V_{gs} - V_{tn}\right)} ln\left(\frac{\Delta V_{logic}}{\Delta V_o}\right) \tag{4.19}$$

where C_L is proportional to the gate-source capacitance of a single transistor ($C_L = K_1 WLC_{ox}$, with K_1 between 1 and 2), $G_m = \frac{A_V}{R_L}$ is the transconductance of each inverter ($G_m = K_2\, gm = K_2\, \mu_n C_{ox} \frac{W}{L} (V_{gs} - V_{tn})$, with K_2 between 0.5 and 1) and $K = \frac{K_1}{K_2}$.

ΔV_o is the initial voltage difference at the beginning of the latch phase and ΔV_{logic} is the voltage difference required in order for the succeeding logic to recognize the correct output.

Equation 4.19 shows a dependence from technology and, more importantly, a dependency from ΔV_o. If the initial voltage difference is small, the time required may be larger then the time allowed for the latch phase, i.e, $T_{latchmin} > T_{latch}$. Such situation is often referred to as *Metastability*.

In digital systems, the *metastable state* refers to the operation of bistable circuits in an unstable equilibrium point under specific input conditions called marginal triggering. Once the bistable circuit enters its metastable region, it can remain in such a state for an indefinite time prior to evolving into one of its stable states. During this time the bistable circuit's output present a voltage value undefined at a logical level. In these conditions, it is impossible to guarantee that two circuits reading the same metastable signal simultaneously, will interpret the same logic value [20].

4.4.3 Offset

The comparator's accuracy depends mainly on how close to the input voltage the nodes V_{o+} and V_{o-} turn out to be at the end of the sensing phase. In practice, these voltages are different due to threshold voltage and size mismatches, charge injection mismatch between input switches (S_1 and S_2 in figure 4.13) and current flowing through switches S_3 and S_4 [21].

This difference is usually referred as *input offset voltage*. To minimize this problem, the Dynamic Latch is preceded by linear preamplifier[10] in order to reduce the offset voltage and achieve the required accuracy. For example, if a maximum input offset voltage of $2\,mV$ is required and the Dynamic Latch has an offset of $4\,mV$, it must be preceded by a linear amplification stage with a voltage gain higher then 2.

[10]Note that the Dynamic Latch is itself an amplifier (non-linear), as it amplifies a small input voltage into the supply rails.

In most cases this need for amplification stage is a serious limitation because it results in higher power consumption. However, for gains around 2 or 3, parametric amplification (discussed at section 4.3.3) may be used with no static power consumption.

Moreover, there is a direct relationship between gain and time-constant in amplifiers, which limits this stage's gain to around 10 ([14]) in high speed solutions.

4.5 Comparator

Perhaps, the most important block in the ADC's architecture (and even more in Flash ADCs) is the comparator. An n-bit Flash ADC requires $2^n - 1$ comparators, each one set to compare the input voltage, V_{in}, with a specific voltage (usually a fraction of an existing reference voltage, V_{ref}). Due to the number of required comparators, most of the power budget is spend in this block. Moreover, these are the main contributors for decreasing the overall ADC's performance.

Before going deeper into the comparator architecture, it is important to remember what was discussed in section 4.3.1 about the equivalent model of a switched capacitor network (4.15).

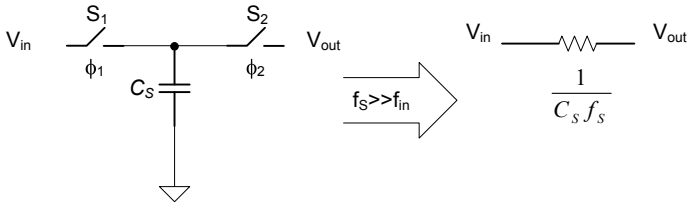


FIGURE 4.15: Switched capacitor equivalent resistor model.

Consider now the circuit in figure 4.16. Here, two switched capacitors are user to achieve an equivalent model with two resistors.

From the equivalent circuit, it is easy to realize that by adjusting the relation between the two sampling capacitors (k_S), the output voltage can be defined as a function of V_{in} (considering V_{ref} and k_S constant):

$$V_{out} = \frac{k_S}{k_S + 1} (V_{in} - V_{ref}) + V_{ref} \tag{4.20}$$

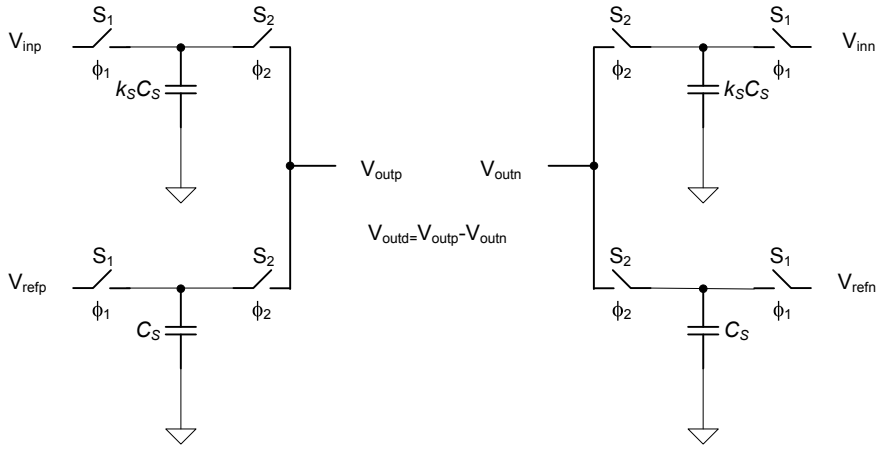


FIGURE 4.16: Switched capacitors as a resistor ladder.

In fact, as will be seen next, this circuit configuration will substitute the resistor ladder in a traditional Flash architecture (see figure 4.2). This approach has two major advantages over the resistor ladder. The first one is that integrated circuit resistors may have an error as high as 25%, which leads to unapropriate reference voltages for the comparators. The second advantage is that the static power consumption in the resistor ladder is eliminated, leading to higher power efficiency.

The comparator block diagram is shown in figure 4.17.

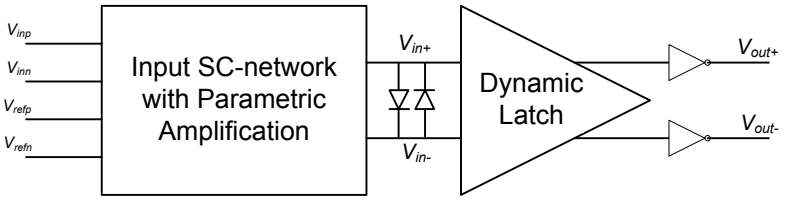


FIGURE 4.17: Comparator block diagram [22].

It is based in an input Switched Capacitor network followed by a dynamic latch proposed by [22]. To overcome the input offset voltage of the latch and minimize metastability problems, parametric amplification is employed in the Switched Capacitor network.

Notice the anti-parallel connected diodes at the input of the dynamic latch. These diodes will limit the latch's output during the latch phase, preventing them to reach the supply voltages. This speeds up the process when the latch is reseted back to the trip-point.

Moreover, a differential structure is used. The main advantages of such structures is well known, and is related with the reduction[11] of second-order distortion (HD_2) and the reduction of the amplified input bias voltage to a common-mode [22]. Additionally, due to the lower supply voltage constrains, the analog input range can be doubled.

In [18], one can find the relation between HD_2 and HD_3 as a function of the input signal amplitude and common-mode. The important point here is that parametric amplification introduces harmonic distortion, thus, differential structure is preferred.

The equivalent model of the circuit in a differential configuration is presented in figure 4.18.

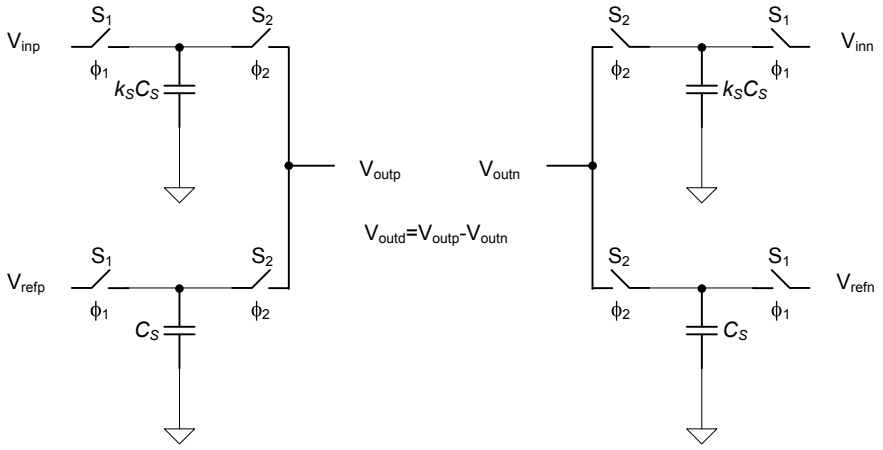


FIGURE 4.18: Switched capacitor in a differential configuration.

For this circuit, equation 4.20 takes the following form:

$$V_{outd} = \frac{k_S}{k_S + 1} \left(V_{ind} - V_{refd} \right) + V_{refd}. \tag{4.21}$$

Including the parametric amplification gain (k_{PA}), discussed in section 4.3.3, one can obtain:

$$V_{outd} = \frac{k_S}{k_S + 1} \left(k_{PA} V_{ind} - k_{PA} V_{refd} \right) + k_{PA} V_{refd}. \tag{4.22}$$

The goal in this work is to implement a 4-bit Flash ADC in CMOS technology. In particular the UMC's 130 nm process, with 1.2 V voltage supply.

[11] Differential structures, in theory, eliminate even-order distortion. However, due to mismatches in the two signal paths, the output has some power present in the ever-order harmonic frequencies of the input signal.

The input differential voltage is defined as $V_{ind} = V_{inp} - V_{inn}$, where V_{inp} and V_{inn} swings between $300\,mV$ and $800\,mV$. This results in a analog input range of $1\,V$ (differential), and a common-mode at the input of $V_{CMi} = 550\,mV$. The reference voltages are $V_{refp} = 800\,mV$ and $V_{refn} = 300\,mV$, resulting in $V_{refd} = 500mV$.

Table 4.1 shows the differential threshold voltage that defines the required comparison level for each comparator. Note that each level is a multiple of the quantization step $\Delta = \frac{1}{2^4} = 62.5\,mV$ (see B.3).

Note the Output column. This will be used when referring to the corresponding comparator's binary output, e.g, the output of comparator C_{125} is $C_{(10)p}$. Since the structure is differential, each comparator also have a complementary output which will conveniently be referred as $C_{(10)n}$, to keep in the same example.

Comparator	Output	Threshold Voltage (mV)
$C_{437.5}$	$C_{(15)p}$	$7/8\,V_{refd} = 437.5$
C_{375}	$C_{(14)p}$	$3/4\,V_{refd} = 375$
$C_{312.5}$	$C_{(13)p}$	$5/8\,V_{refd} = 312.5$
C_{250}	$C_{(12)p}$	$1/2\,V_{refd} = 250$
$C_{187.5}$	$C_{(11)p}$	$3/8\,V_{refd} = 187.5$
C_{125}	$C_{(10)p}$	$1/4\,V_{refd} = 125$
$C_{62.5}$	$C_{(9)p}$	$1/8\,V_{refd} = 62.5$
C_0	$C_{(8)p}$	0
$C_{-62.5}$	$C_{(7)p}$	$-1/8\,V_{refd} = -62.5$
C_{-125}	$C_{(6)p}$	$-1/4\,V_{refd} = -125$
$C_{-187.5}$	$C_{(5)p}$	$-3/8\,V_{refd} = -187.5$
C_{-250}	$C_{(4)p}$	$-1/2\,V_{refd} = -250$
$C_{-312.5}$	$C_{(3)p}$	$-5/8\,V_{refd} = -312.5$
C_{-375}	$C_{(2)p}$	$-3/4\,V_{refd} = -375$
$C_{-437.5}$	$C_{(1)p}$	$-7/8\,V_{refd} = -437.5$

TABLE 4.1: Ideal threshold differential voltages of the 15 comparators in a 4-bit ADC

In this work, particular attention was be given to the comparator used for both $+125\,mV$ and $-125\,mV$ threshold levels. In fact, these two are exactly equal except in the input voltage sources, i.e, the V_{inp} source is connected to the V_{inn} input of the comparator to achieve a $+125\,mV$ comparator from a C_{-125} one. The remaining ones are easily dimensioned based on the analysis of these two.

4.6 Conclusion

The complete comparator's schematic is presented in figure 4.19. It was proposed by [22] and this work was a follow up of that idea. The shaded areas highlight the project's most relevant structures. A more detailed analysis is done in section 5.3.

Note that the circuit is an almost perfect mirror image in respect to the center. This is to allow differential operation, i.e., the V_{inn} path must be as close as possible to the V_{inp}[12].

[12]Note that this circuit is not fully differential because there is no common mode feedback circuit linking both branches, the mirror configuration achieves, to some extend, many of the fully differential benefits. This configuration is usually referred as *pseudo-differential*.

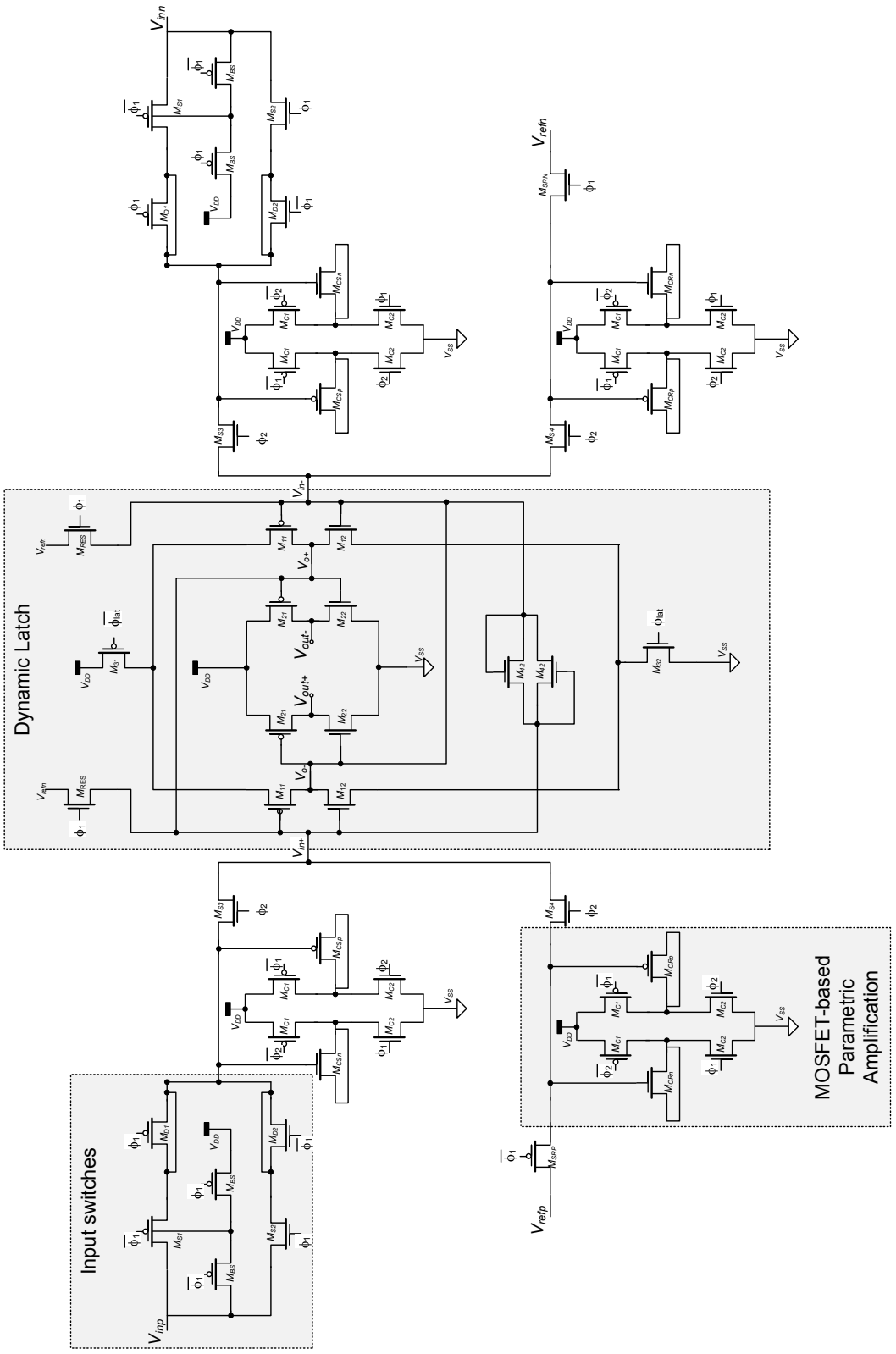


FIGURE 4.19: Complete comparator's schematic.

Chapter 5

Implementation of a Flash ADC With Parametric Amplification

5.1 Introdution

The results from the previous chapters are used to implement a Flash ADC in a 130 nm CMOS technology. This chapter presents the dimensioning process as well as simulation results obtained.

5.2 Proposed Architecture

Figure 5.1 illustrates the top level architecture used. This can be related with the initial block diagram of an ADC presented in section 4.1, figure 4.1. The *Sampling* and *Quantizer* functions are realized by the comparator bank, and the bubble detector together with the ROM codifier realize de *Codifier* block.

The inputs/outputs of each block, as well as the clock scheme used is also presented. Note that the ϕ_2 is an inverted version of ϕ_1, meaning that no non-overlap clock was used.

The following sections deal with each block separately.

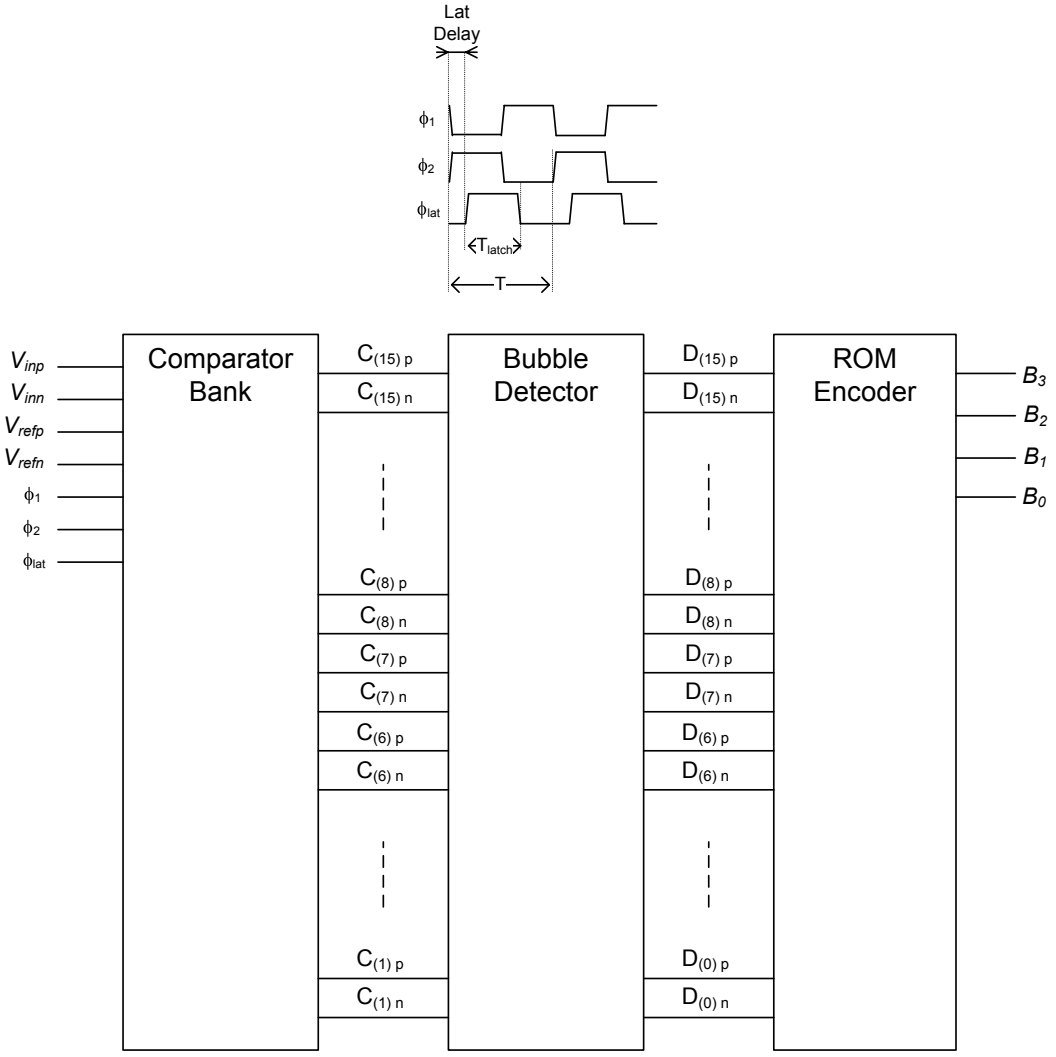


FIGURE 5.1: Proposed Architecture.

5.2.1 Comparator's Bank

The comparator bank is composed by $2^4 - 1 = 15$ comparators, each one set to compare the input voltage to a different reference voltage (see table 4.1). Figure 5.2 illustrates this by stacking the comparators o top of each other, starting with the lowest threshold comparator.

The simple observation that, for a given $V_{ind} = V_{inp} - V_{inn} = V_y$, one can obtain $V_{ind} = -V_y$ just by swapping the inputs, as mentioned in section 4.5, results in the cross wires in the signal path. This means that only 8 different comparators are needed.

The comparator's bank output is usually referred as a *Thermometer code*. This is mainly because the output changes sequentially as the input voltage rises. Table 5.1 illustrates

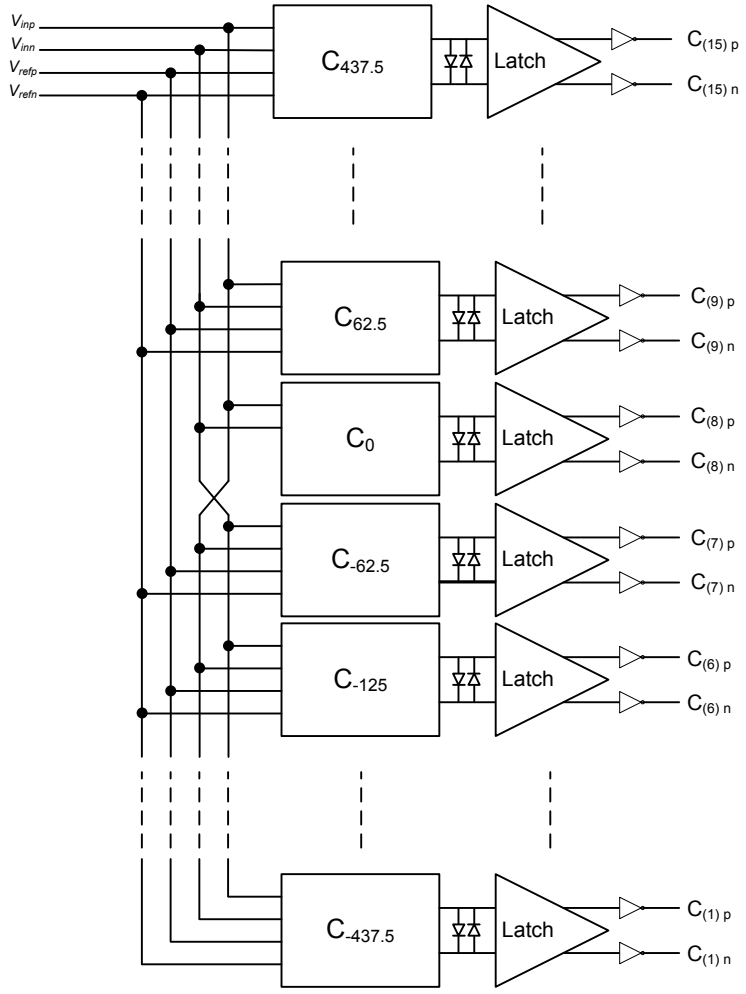


FIGURE 5.2: Comparator Bank.

the output state of each comparator for different input voltages.

The thermometer's output of the comparator's bank is not a binary code as wanted. To transform this thermometer code into a binary code, one can use a *bubble detector*, followed by an *ROM* codifier, as will be explain next.

5.2.2 Bubble Detector

This block, together with the ROM codifier, is responsible for the transformation of the comparator's bank output into a binary code. These two blocks realize the *Codifier* block in the ADC's block diagram presented in figure 4.1. The odd name "bubble detector" came from a additional feature and will be explain shortly.

	V_{ind}			
Comparator	$-450\,mV$	$-400\,mV$	$30\,mV$	$200\,mV$
$C_{(15)p}$	0	0	0	0
$C_{(14)p}$	0	0	0	0
$C_{(13)p}$	0	0	0	0
$C_{(12)p}$	0	0	0	0
$C_{(11)p}$	0	0	0	1
$C_{(10)p}$	0	0	0	1
$C_{(9)p}$	0	0	0	1
$C_{(8)p}$	0	0	1	1
$C_{(7)p}$	0	0	1	1
$C_{(6)p}$	0	0	1	1
$C_{(5)p}$	0	0	1	1
$C_{(4)p}$	0	0	1	1
$C_{(3)p}$	0	0	1	1
$C_{(2)p}$	0	0	1	1
$C_{(1)p}$	0	1	1	1

TABLE 5.1: Thermometer code

The goal here is to transform the thermometer code into a code containing *only* one 1 in it's output, as illustrated in table 5.2. The easiest way to realize this function is with digital logic and, in the ideal case, a *NAND* gate with two inputs is enough to detect the $0 \rightarrow 1$ transition in the thermometer code.

Connecting two consecutive comparators to a NAND gate realizes this function, as figure 5.3 illustrates. Note that any different input results in a 0 output.

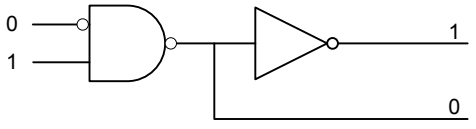


FIGURE 5.3: NAND gate to detect the $0 \rightarrow 1$ transition.

The problem with this approach is the assumption of ideal operation in all the comparators. As will be seen next, the ROM codifier requires a single 1 at the input to work properly. But if, for some reason, one of the comparators output is 0 when it should be 1 (or the other way around), there will be two $0 \rightarrow 1$ detections, the correct one, and the "stuck" comparator's one. To enter into account for this possibility one can use a *Bubble Detector*. This circuits detect 0s (or bubbles) below the thermometer's top level and discard them.

Comparator	Thermometer code	$0 \rightarrow 1$ Detector
$C_{(15)p}$	0	0
$C_{(14)p}$	0	0
$C_{(13)p}$	0	0
$C_{(12)p}$	0	0
$C_{(11)p}$	0	0
$C_{(10)p}$	0	0
$C_{(9)p}$	0	0
$C_{(8)p}$	1	1
$C_{(7)p}$	1	0
$C_{(6)p}$	1	0
$C_{(5)p}$	1	0
$C_{(4)p}$	1	0
$C_{(3)p}$	1	0
$C_{(2)p}$	1	0
$C_{(1)p}$	1	0

TABLE 5.2: $0 \rightarrow 1$ detection

In this work, a first order error correction is used, meaning that the bubble detector can resolve multiple but isolated 0s. Other more complex circuits can eliminate two, three or more.

Figure 5.4 shows a part of the bubble detector's circuit. The complete circuit can be easily extrapolated.

With this configuration, instead of detecting the $0 \rightarrow 1$ transition, it detects the "001" sequence at the comparators bank output. As an example of the error correction of this configuration consider the thermometer sequence "000101101". A $0 \rightarrow 1$ detection will find 3 transitions, while the bubble detector finds only one.

To illustrate the bubble detection operation, figure 5.4 presents a possible comparator's bank output with $C_{(7)p}$ "stuck" at 0, thus producing a bubble in the thermometer code. In this situation, the correct output is produced. Note that the problem remains and this malfunction will result in a missing code. However, a missing code is preferred to a complete failure.

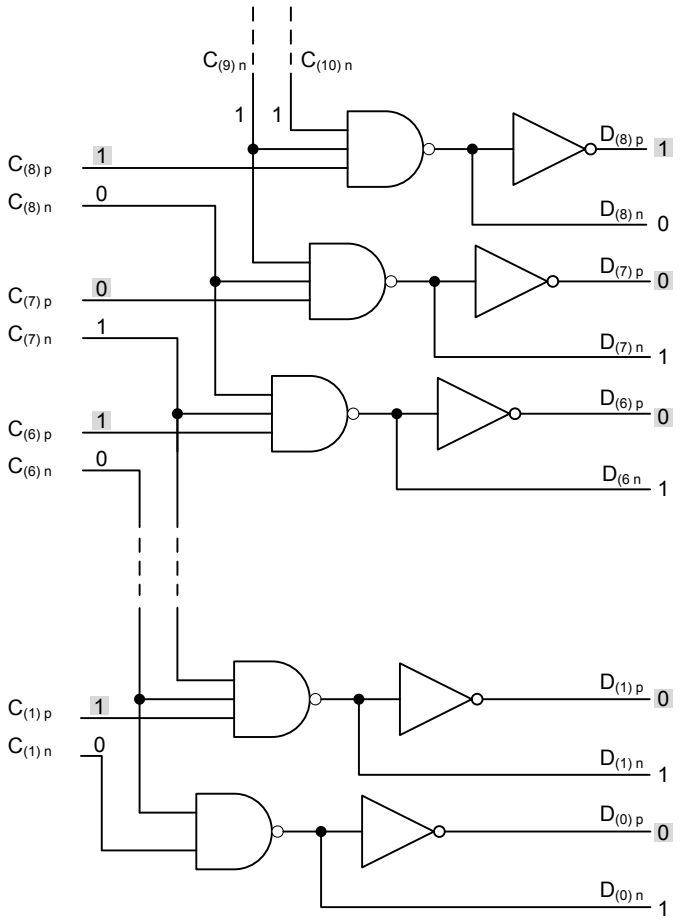


FIGURE 5.4: Bubble detector schematic.

5.2.3 ROM Encoder

The second step to codify the thermometer code to a binary word is the ROM codifier, where the name "ROM" comes from the well known read-only memory from digital systems. The principle here is very simple and can easily understood by the example schematic in figure 5.5.

If $D_{(x)p} = 1$ and $D_{(x)n} = 0$, all the transistors act as *on*-switches, and the produced output code in this example is "1000" ($B_3B_2B_1B_0$). To keep the analogy with digital circuits, the bubble detector act as an address pointer, and the ROM codifier stores an output code for each address. By rearranging the transistors' configuration, each address ($D_{(x)p/n}, 1 \leq x \leq 15$) generate a different code.

In this work, for simplicity and ease in simulation results analysis, the output is a binary

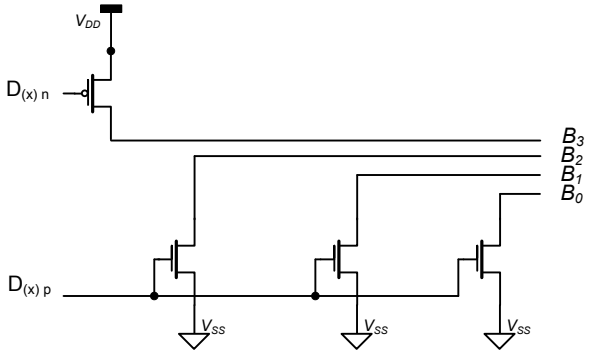


FIGURE 5.5: ROM example schematic (one address line).

code. However, the reflected binary code (or Gray code) offers some forward error correction possibilities (beyond the scope of this work) that make it more efficient. The change between binary code and gray code can be achieved simply by rearranging the transistors configuration for each address.

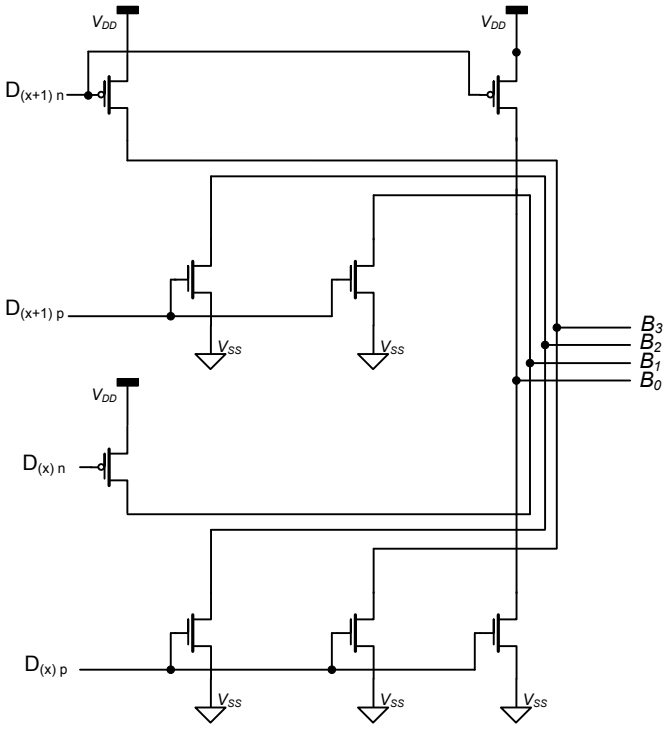


FIGURE 5.6: ROM example with two address lines.

Figure 5.6 gives some more insight on how the ROM quantifier complete schematic should be. In this example, the output code is "1000" for the address $D_{(x)}$ and "1001" for the $D_{(x+1)}$. Note that, if the bubble detector produces both the $D_{(x)}$ and $D_{(x+1)}$ outputs simultaneous (which is a non-ideal but possible behavior), the least significant bit (B_0, corresponding to the changing bit from one code to another) is undefined because both

the NMOS and PMOS transistors are *on*. This behavior reinforces the need for bubble detection discussed above.

5.3 Design Methodology

Nowadays, computers play a major role in almost every tasks in the scientific world, and integrated circuit design is no exception. Computer-Aided Design (CAD) tools, together with accurate component models, allows designers to quickly simulate circuits using complex models, which would be almost impossible to achieve with just a pen and a note book.

As it turns out, the first or second order approaches usually used when analyzing a circuit are, in many cases, inaccurate enough to result in disappointing results after implementation. As IC technology evolves and operating frequencies increases, parasitic capacitances, short channel effects, electrical phenomena and others, forces the designer to rely in simulations, rather than restrict dimensioning to theoretic expressions.

However, theoretical analysis also plays a central role, as it allows the designer to achieve initial dimensioning and infer probable causes for unexpected simulation results, as well as decide in which direction to take in order to make the circuit work and achieve the best performance possible.

Moreover, some degree of experience and sensibility (many times obtained by trial-and-error) is essential. Knowing ahead what works and what doesn't, may be the best starting point when experimenting new ideas. For this reasons, a complete description of the dimensioning process for each and every one of the transistors in the circuit is not the goal here and would be tedious. The dimensioning workflow will be restricted to the most relevant parts of the circuit.

For the following dimensioning process, results and analysis from previous sections will be used as a theoretical support.

5.3.1 Comparator

As already mentioned, the starting point for this project was the work presented in [22]. To start dimensioning, one must start by the center piece of the overall architecture, in

this case, the comparator. The following text will focus on the three shaded structures in the comparator's schematic presented in figure 4.19.

Ideally, a comparator switches exactly when one input is infinitesimally larger than the other. If so, there is no need for gain stages and the input signal could be sampled and applied directly to the dynamic latch. In practice, the input offset voltage determines the minimum amount of gain needed to ensure accurate comparison.

Dynamic Latch

The design procedure should start with by sizing the dynamic latch and output buffers transistors because the parametric amplification gain is degraded by these input capacitance and the comparator's accuracy is determined by the dynamic latch's input offset.

In the dynamic latch's schematic presented in figure 5.7, transistors M_{11} and M_{12}[1] form the two back-to-back inverters. M_{31} and M_{32} turn on and off the positive feedback regeneration and the diode connected transistors M_{41} and M_{42} limit the output. Switches M_{res} reset the latch during the sample phase to eliminate memory effect and, finally, the output is buffered by transistors M_{21} and M_{22} in an inverter configuration.

The main considerations to account for the dynamic latch are the input capacitance, the input referred offset and the time constant. To reduce the time constant, L_{min} is used, as equation 4.19 suggests. Also, to achieve low input capacitance, transistors M_{11}, M_{12}, M_{21} and M_{22} were kept small. Finally, to allow fast regeneration during the latch phase, transistors M_{31} and M_{32} were over dimensioned.

Moreover, due to the relation between electrons and holes mobility, the PMOS transistors are usually 3 or 4 times wider than NMOS[2] to reach similar transconductance factor.

According to [23], the transistor's dimensions and threshold voltages mismatches contribution to the offset voltage in a CMOS Latch can be obtained from:

$$V_{os} = \frac{V_X - V_{tn}}{2\left(1+\sqrt{\frac{k_p}{k_n}}\right)}\frac{\Delta k_p}{k_p} - \frac{V_X - V_{tn}}{2\left(1+\sqrt{\frac{k_p}{k_n}}\right)}\frac{\Delta k_n}{k_n} - \frac{\sqrt{\frac{k_p}{k_n}}}{1+\sqrt{\frac{k_p}{k_n}}}\Delta V_{tp} + \frac{1}{1+\sqrt{\frac{k_p}{k_n}}}\Delta V_{tn} \quad (5.1)$$

[1]Note that there are more than one transistors with the same label. This indicates that their dimensions will be the same.

[2]This is the same as in the transmission gate switches, the actual value should be $\frac{\mu_n}{\mu_p}$.

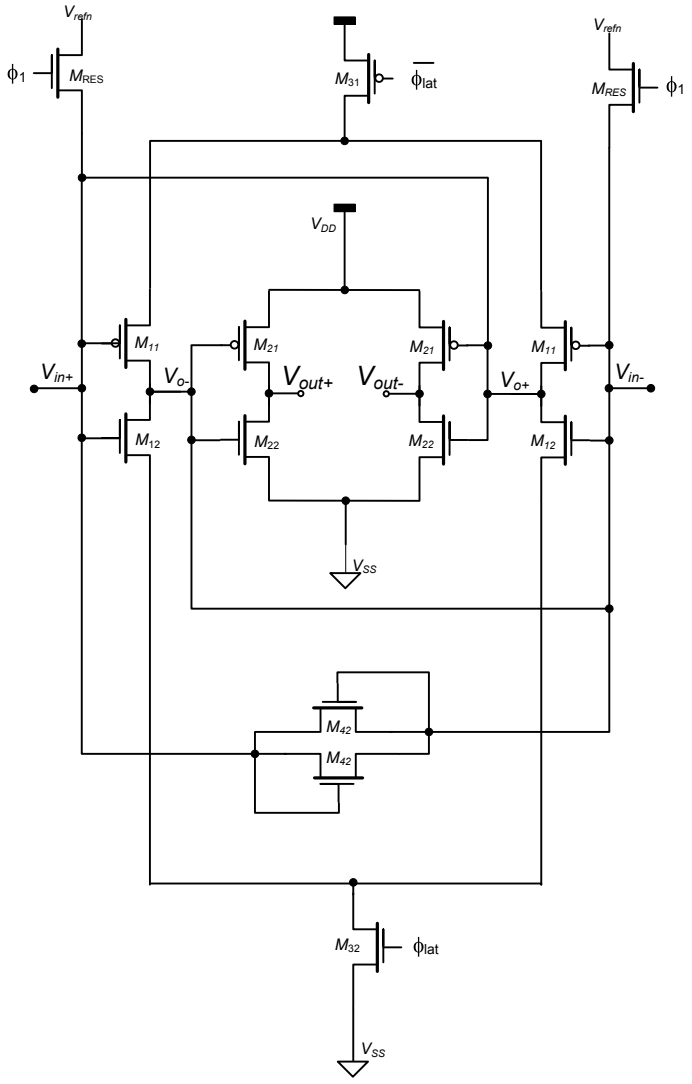


FIGURE 5.7: Dynamic Latch schematic.

where

$$V_X = \frac{\sqrt{\frac{k_p}{k_n}}\,(V_{DD} - V_{tp}) + V_{tn}}{1 + \sqrt{\frac{k_p}{k_n}}} \tag{5.2}$$

is the input voltage of an inverter that causes the vertical linear section in the transfer function, as illustrated in figure 5.8. V_{tp} and V_{tn} are the threshold voltages of PMOS and NMOS transistors and $k_{p/n} = \frac{\mu_{p/n} C_{ox}}{2} \frac{W_{p/n}}{L_{p/n}}$.

Note that, if $\frac{W_p}{W_n} = \frac{\mu_n}{\mu_p}$ equation 5.2 results in $V_X = V_{DD}/2$.

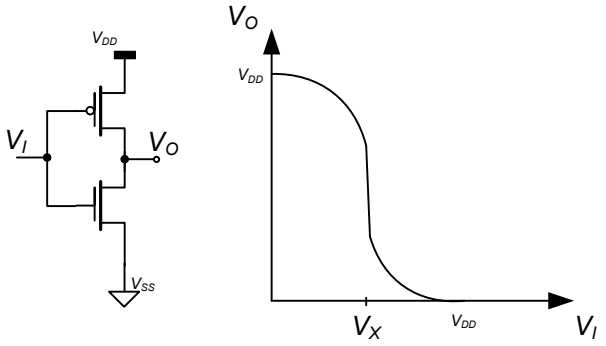


FIGURE 5.8: CMOS inverter and Transfer Function.

With $M_{11} = 8\mu m/0.13\mu m(W/L)$ and $M_{12} = 2\mu m/0.13\mu m(W/L)$, this offset is $7mV$ for the deviation parameters of the UMC 130nm technology. This only enters into account for the dimension and threshold voltage mismatches. Adding the contribution of charge injection mismatch between the switches at the latch input, determines the latch input resolution, which was set to $40mV$ [22].

Table 5.3 shows the result of this dimensioning process.

Transistor	W	L
M_{11}	$8\mu m$	$0.13\mu m$
M_{12}	$2\mu m$	$0.13\mu m$
M_{21}	$2\mu m$	$0.13\mu m$
M_{22}	$0.5\mu m$	$0.13\mu m$
M_{31}	$36\mu m$	$0.13\mu m$
M_{32}	$12\mu m$	$0.13\mu m$
M_{42}	$0.25\mu m$	$0.13\mu m$
M_{res}	$1\mu m$	$0.13\mu m$

TABLE 5.3: Dimensions for the Dynamic Latch transistors.

Figure 5.9 shows the simulation result of a dynamic latch. In this, the input differential voltage is set to be a very slow ramp, ranging from $50\,mV$ to $-50\,mV$ and then back to $50\,mV$. This way, one can verify the latch's trip-point (that should be when the input differential voltage crosses zero) and some memory effect.

It comes with no surprise that the offset value obtained with this simulation is negligible. In fact, this simulation does not enter into account with any process mismatch, thus, the $\pm 100\mu V$ offset for the ideal trip-point can be attributed to the fact that the latch is in the latch phase when the input crosses zero. The new decision will only occur in the next

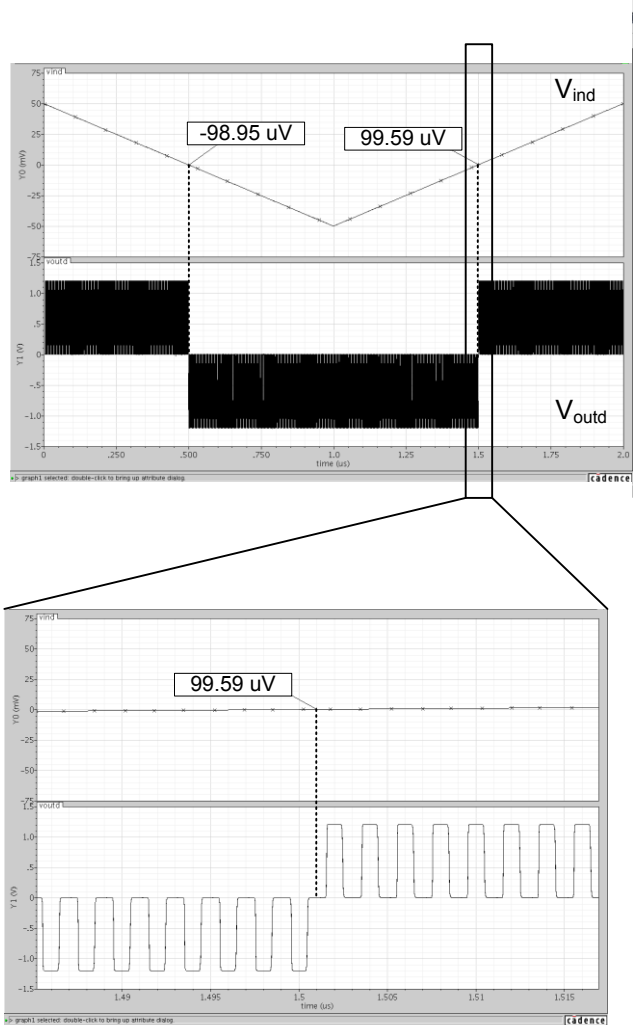


FIGURE 5.9: Dynamic Latch Hysteresis electrical simulation and analysis.

latch phase, when the input voltage is higher (or lower, depending on the input voltage ramp).

The most important result here is that the memory effect, which would result in a hysteresis effect, appears to have been eliminated with the reset operation during the sample phase (transistors M_{res}).

C125mV MOSFET Sampling Capacitors

An important parameter that must be defined before going any further is the comparator's accuracy. Remember that the quantization step for this ADC is $\Delta = \frac{1}{2^4} = 62.5mV$. This means that each comparator must have a maximum offset of $\pm\frac{1}{2}LSB = \frac{1}{2^{4+1}}$. To give

some headroom for unaccounted parameters, a maximum offset of $\frac{1}{2^{4+2}} = 15.625mV$ is set.

The sizing of the MOSFET sampling capacitor (MOScap) starts by analyzing the gain requirements imposed by the dynamic latch's input offset voltage. The need for a pre-amplification stage is related to the need to reduce this unavoidable offset to reasonable values. Considering a comparison level of $\frac{V_{refd}}{4} = -125mV$ and an input latch capacitance, $C_L = 50fF$, from [22] a pre-amp gain of 2.5 can be achieved by squared transistors with $WL = 2 \times 2\mu m^2$ and $1 \times 1\mu m^2$, respectively for $M_{CSn/p}$ and $M_{CRn/p}$ in figure 4.19.

Care must be taken in not using either large L (due to speed limitations), or minimum L (due to short-length effects). Moreover, a larger L should results in added difficult when removing the inversion charges during the boost phase. This last consideration indicated that the length should be reduced. As a consequence, this modified structure with $M_{CSn/p} = 4 \times 0.5\mu m^2$ and $M_{CRn/p} = 1 \times 0.5\mu m^2$, also decrease the effect of the extrinsic (overlap) gate capacitances during the amplification phase, increasing the gain.

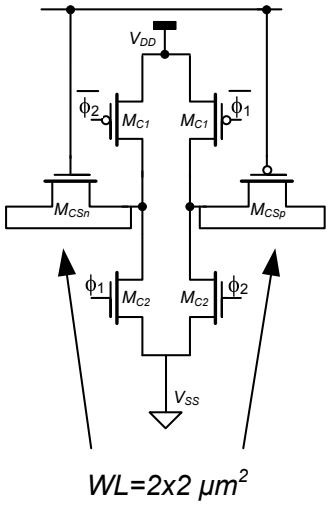
To achieve the same capacitance value, a par of each (M_{CSn} and M_{CSp}) must be used, as illustrated in figure 5.10(b). The same applies to the reference sampling capacitors, except in these, the width was doubled to achieve the same capacitance[3].

Note that the area (WL) of $M_{CSn/p}$ is 4 times the one of $M_{CRn/p}$. This fact is related to the comparison level $1/4\, V_{refd} = 125mV$. Thus, as already mentioned, the remaining comparators can be easily sized just by adjusting the $\frac{M_{CSn/p}}{M_{CRn/p}}$ to the desired comparison level.

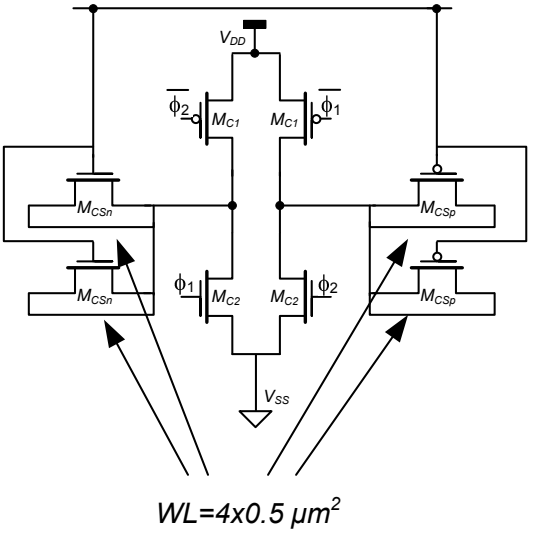
This result is also supported by equation 4.21, if one considers an offset-free dynamic latch (thus, pre-amplification is not necessary), i.e., $V_{outd} = 0$ for $k_S = 4$, $V_{ind} = -125mV$ and $V_{refd} = 500mV$, as expected.

The only special consideration is the C_0 comparator. For this, the reference branches are simply removed from the C_{125} comparator (but could be from any other). This way, the dynamic latch trip-point is when both inputs are equal, or $V_{ind} = 0$.

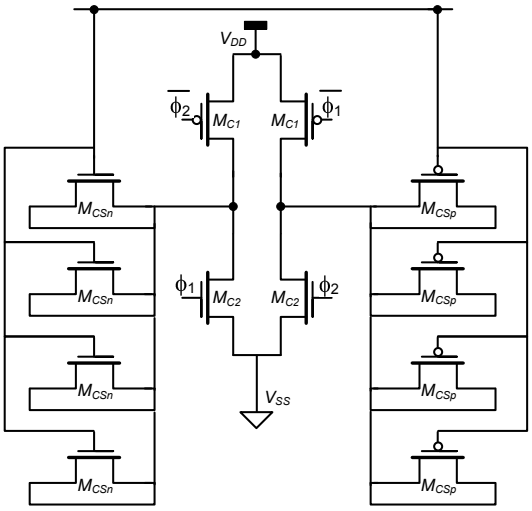
[3]Both doubling the width or the number of transistors results in doubling the area. Due to layout disposition, a maximum width of $4\mu m$ is preferred because *Fingers* will not be used to allow easy access to the source and drain terminal. Moreover, smaller transistors allow smaller substrate resistance noise [17] if the transistor is surrounded by a guard-ring connecting the subtract to V_{SS} or V_{DD} (NMOS or PMOS).



(a) MOScaps with $2 \times \left(2 \times 2\mu m^2\right)$ (total area = $8\mu m^2$)



(b) MOScaps with $4 \times \left(4 \times 0.5\mu m^2\right)$ (total area = $8\mu m^2$)



(c) MOScaps with $8 \times \left(4 \times 0.5\mu m^2\right)$ (total area = $16\mu m^2$)

FIGURE 5.10: Parametric MOScaps

The simulated results for a single NMOS transistor indicate that the total gate capacitance of a $4 \times 0.5\mu m^2$ is $19fF$, while the theoretical value ($C_g \approx WLC'_{ox}$) is 25fF. By assuming equal gate capacitance for the PMOS transistor, the total sampling capacitance for the two pair of transistors is assumed to be $2 \times 50 = 100fF$ (worst case for time constant analyses and best case for kT/C noise).

Concerning the kT/C noise, the total power stored in each sampling capacitor is $P_{n,C_s} = 41.4nW$ at room temperature (B.4) or $v_{n,C_s} = 203.5\mu V$. Comparing this with the quantization noise, $V_{\epsilon Q} = V_{LSB}/\sqrt{12} = 44mV$ allow the consideration that the kT/C is insignificant, and the used sampling capacitances are big enough. However, the $40mV$ input offset of the dynamic latch is reduced to $16mV$ at the comparator's input for a gain of 2.5. This value is bigger than the desired $15.625mV$.

To comply with the specifications, the parametric amplification must have a gain of, at least, 4. As discussed in section 4.3.3 this can be achieved by increasing the sampling capacitance which, in this case, leads to an increase in the total transistors' area. The result is shown in figure 5.10(c) and the final sizes for the MOScap transistors in the C_{125mV} comparator are shown in table 5.4

Transistor	W	L
$4 \times M_{CSn}$	$4\mu m$	$0.5\mu m$
$4 \times M_{CSp}$	$4\mu m$	$0.5\mu m$
M_{CRn}	$4\mu m$	$0.5\mu m$
M_{CRp}	$4\mu m$	$0.5\mu m$

TABLE 5.4: Dimensions for the MOScap transistors in the C_{125mV} comparator.

Input Switches

The input switches turn out to be one of the most important elements when defining the comparator's accuracy if the gain requirements are fulfilled.

A special attention was given to the signal's input switches, shown in figure 5.11.

Note that this switch include a transmission gate (transistors M_{S1} and M_{S2}) with bulk switching in the PMOS transistor (realized by transistors M_{BS}) and also two dummy switches to reduce charge injection (M_{D1} and M_{D2}). The dimensioning process starts by sizing just the transmission gate pair.

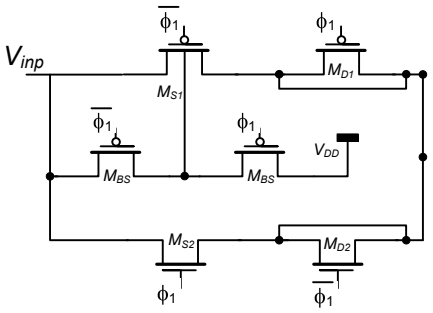


FIGURE 5.11: Input switches' schematic.

From the well known capacitor's charge equation, $V_O = V_I\left(1 - e^{-\frac{t}{\tau}}\right)$, to achieve a minimum of $2mV$[4] accuracy for a $800mV$ input voltage (V_{inp_max}) one must have $t = T_{sampling} \geq 6\tau$, at least.

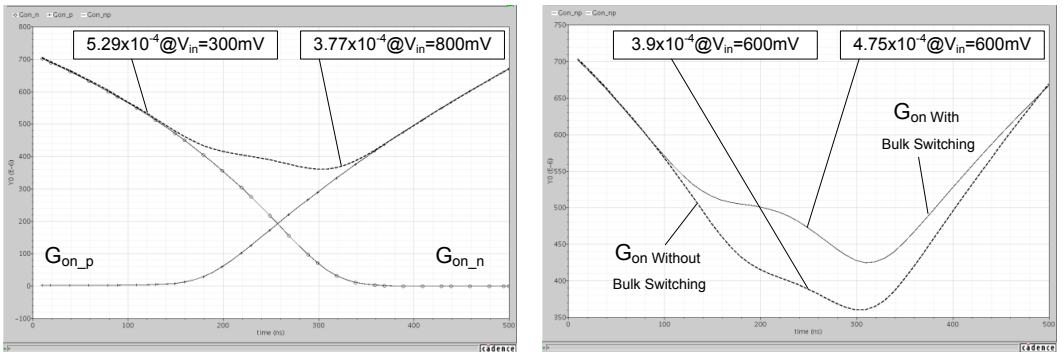
After the considerations made for the sampling capacitor, the input switches must be sized in order to comply with the time constant ($\tau = RC$) allowed for a given frequency. To start by aiming high, a sampling frequency of $f_s = 500MHz$ was considered. Thus, the time for the sampling phase is $T_{sampling} = \frac{1}{2f_s} = 1ns$.

Recalling that the dummy switches should have half the width of the actual switches, that PMOS transistors must be 4 times wider than NMOS and, for speed purposes, minimum length should be used, the input switch sizing starts by using minimum width ($0.16\mu m$) in the dummy and bulk-switching switches, mainly for area minimization. This leads to $M_{S1} = 1.28\mu m/0.13\mu m(W/L)$ and $M_{S2} = 0.32\mu m/0.13\mu m(W/L)$.

Following equation 4.6 and using the transmission gate *on*-conductance result ($G_{on} = G_{on_p} + G_{on_n}$), one can find the value of the input switch conductance to be $0.67mS$ when $300mV \leq V_{in} \leq 800mV$.

Figure 5.12(a) shows the simulated conductance values obtained for PMOS and NMOS transistors with the mentioned dimensions. The dashed line refers to the transmission gate conductance. Due to second order effects and body effects, not accounted in equation 4.6, the flat region illustrated in figure 4.6(c), is not flat at all and the theoretical values are slightly different from the simulated ones. Some improvements can be achieve by the bulk switching technique, as illustrated in figure 5.12(b).

[4]This are $2mV$ out of the $15.625mV$ already defined as maximum offset. The rest is left for other error sources, such as the process mismatches and charge injection that lead to input offset voltage in the dynamic latch.



(a) G_{on} of NMOS and PMOS transistors

(b) Bulk switching effect

FIGURE 5.12: *On*-conductance simulation results.

Using simulation results, rather that theoretical ones, for $G_{on} = 450mS$ and $C = 50fF$, the required time to achieve the defined accuracy of $2mV$ is $6\,\tau \approx 0.7ns$, which is lower than $T_{sampling}$, as wanted.

To follow up with the increased sampling capacitance described in the previous section, the input switch must increase its conductance by 4, setting $M_{S1} = 5.12\mu m/0.13\mu m(W/L)$ and $M_{S2} = 1.28\mu m/0.13\mu m(W/L)$. To account for parasitic capacitances during the sample phase, the final values were rounded up for $M_{S1} = 8\mu m/0.13\mu m(W/L)$ and $M_{S2} = 2\mu m/0.13\mu m(W/L)$.

Table 5.5 shows the final sizing for the transistors in figure 5.11.

Transistor	W	L
M_{S1}	$8\mu m$	$0.13\mu m$
M_{S2}	$2\mu m$	$0.13\mu m$
M_{D1}	$4\mu m$	$0.13\mu m$
M_{D2}	$1\mu m$	$0.13\mu m$
M_{BS}	$1\mu m$	$0.13\mu m$

TABLE 5.5: Dimensions for the input switch transistors

Although the input switches are implemented with this circuit, the remaining ones use only single NMOS or PMOS devices, depending on the expected input voltage. Transistors M_{C1} and M_{C2} were made big to facilitate the removing of the inversion charges from

the channel. M_{S3} is in the signal path, so it also should have low resistance. Table 5.6 summarizes these transistors dimensions.

Transistor	W	L
M_{S3}	$5\mu m$	$0.13\mu m$
M_{S4}	$2\mu m$	$0.13\mu m$
M_{SRp}	$2\mu m$	$0.13\mu m$
M_{SRn}	$2\mu m$	$0.13\mu m$
M_{C1}	$12\mu m$	$0.13\mu m$
M_{C2}	$3\mu m$	$0.13\mu m$

TABLE 5.6: Sizing for the remaining input switches

5.3.2 Bubble Detector

The bubble detector block is composed only by NAND gates and inverters. It is common practice to keep all digital logic gates with equal dimensions or, at most, double them from stage to stage. This is done to keep the input capacitances at reasonable values and avoid slow rise/fall times in digital logic. Using minimum length minimizes input capacitance and increases speed.

Moreover, the PMOS transistors were set to be 4 times wider than NMOS to achieve similar transconductance factor. This also sets the logic "decision point" (the input voltage that leads to a 0 or 1 logic value at the output, illustrated in figure 5.8) around $V_{DD}/2$.

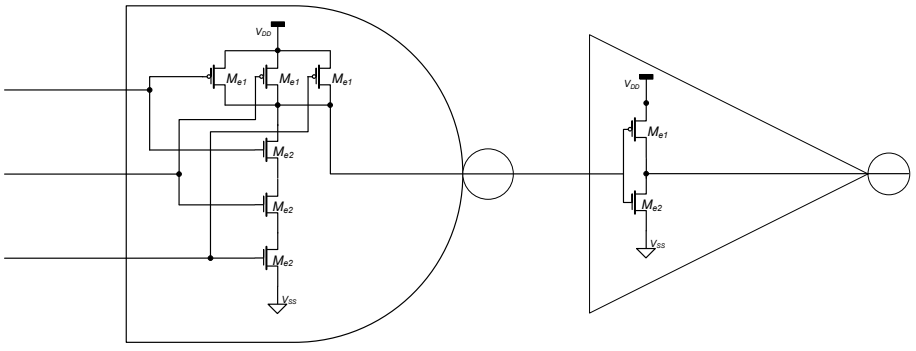


FIGURE 5.13: AND gate.

Figure 5.13 shows the NAND circuit implementation, followed by an inverter to achieve an AND gate. Note that both the AND and the NAND outputs are available. This will be useful for the ROM codifier.

Recalling that the output buffers from the Latch were $M_{21} = 2\mu m/0.13\mu m(W/L)$ and $M_{22} = 0.5\mu m/0.13\mu m(W/L)$, the rule indicates that the following logic should keep the same dimensions. However, analyses of figure 5.4 indicates that the negative output of the comparator is connected to two different AND gates, so the decision here was to reduce the dimensions to half in order to keep this logic block running as fast as possible and minimize different propagation times.

The final dimensions are indicated in table 5.7.

Transistor	W	L
M_{e1}	$1\mu m$	$0.13\mu m$
M_{e2}	$0.25\mu m$	$0.13\mu m$

TABLE 5.7: Dimensions used in the AND gate transistors

5.3.3 ROM codifier

The same consideration of size were taken into account for the ROM transistors, i.e., $W = 1\mu m$ for the PMOS transistors and $W = 0.25\mu m$ for NMOS, all with minimum length for speed reasons.

5.4 Simulation Results for 4 bit Flash ADC

To validate the circuit design and dimensioning, an endless number of step-by-step simulations have been analyzed. The most relevant ones will be presented in this section. The initial simulations intend to validate the C_{-125mV} comparator and analyze its performance with typical process parameters at $500MHz$. The clock scheme used is presented in figure 5.14.

The time required for charge distribution and parametric amplification (lat_delay) is yet to be defined. In [22] the indicated value of $60ps$ can only be used as a reference due to the several differences in this work.

The simulation result shown in figure 5.15 is a parametric simulation for different delays in the latch phase. It shows the threshold voltage for the comparator reaches constant values for delays above $120ps$.

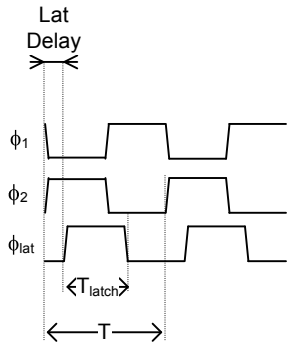


FIGURE 5.14: Clock scheme

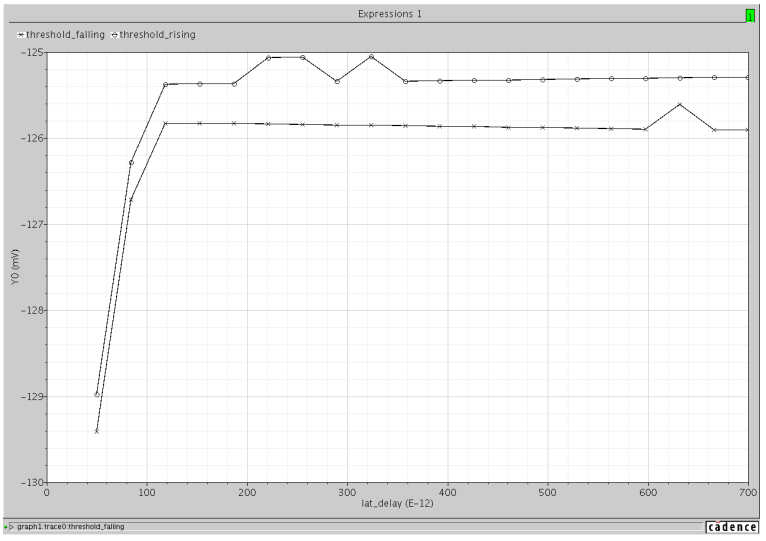


FIGURE 5.15: Simulated comparator's threshold voltages *vs.* Lat_delay parameter.

The indicated threshold_falling and threshold_rising refer to the achieved comparators input differential voltage when the output switches from one state to the other. Due to some hysteresis behavior, this value depends from the previous state. So, the _falling indicates that the input is a ramp with negative derivative and the _rising is the opposite. A delay of $200ps$ was set, and the output result is shown in figure 5.16.

At this point, the hysteresis behavior needs some attention. With typical values for the transistors assumed for simulation, this can not be attributed to the dynamic latch, as discussed in section 4.4. The most probable cause is the time allowed for the sampling period. Note that, depending on the previous state, the sampling capacitor needs to be charge or discharged to the input voltage, resulting in different sampled values, as illustrated in figure 5.17.

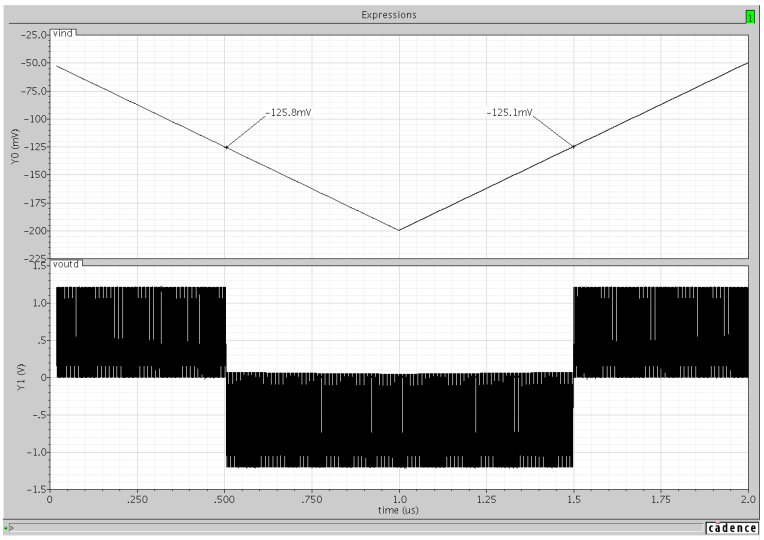


FIGURE 5.16: Simulated comparator's differential input and output.

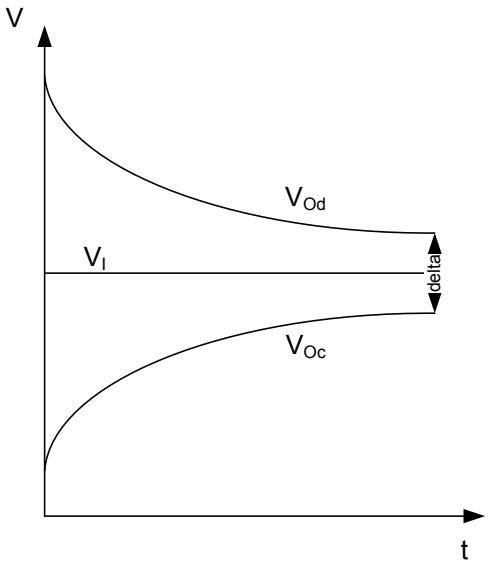


FIGURE 5.17: Charge and discharge of a capacitor.

To analyze this, a parametric simulation can be used. Changing the width of the transistors at the input switches changes the time constant of the RC network and, in turns, changes the *delta* amplitude indicated in figure 5.17. The result is shown in figure 5.18 where the W_{swi} parameter is a fraction of the defined widths for the input switches' transistors and the simulation respects the considerations previously made (e.g. PMOS 4 times wider than NMOS).

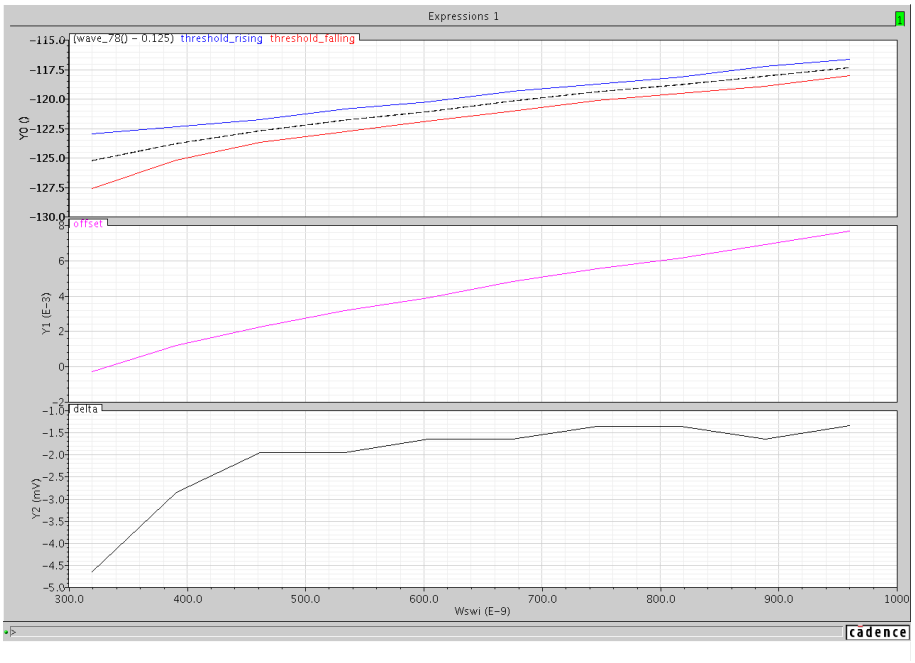


FIGURE 5.18: Simulated Hysteresis analyses.

As expected, for $W_{swi} = 500\,nm$ (which results in the dimensions found in table 5.5) the time constant is less than the sampling time and difference between the two thresholds reaches a stable value, below that the *delta* value is increased. To achieve $delta = 0$ an infinite amount of time would be needed.

Also, note the offset parameter. This indicates the deviation between the wanted $-125mV$ threshold and the achieved average, $\frac{\text{Threshold\_falling+Threshold\_rising}}{2}$. This is not a critical parameter because it can (and will) be adjusted by minor changes in the $\frac{M_{CSn/p}}{M_{CRn/p}}$ relation. However, the fact that it changes with the increase of the input switches indicate some degree of sensibility to parasitic capacitances in the sampling nodes, i.e., wider transistors result in better accuracy but the increase in parasitics decreases the parametric amplification gain and changes the $\frac{M_{CSn/p}}{M_{CRn/p}}$ balance.

To end this simulations set, some changes were made in order to analyze the possibility to achieve higher sampling frequency. The result is shown in figure 5.19.

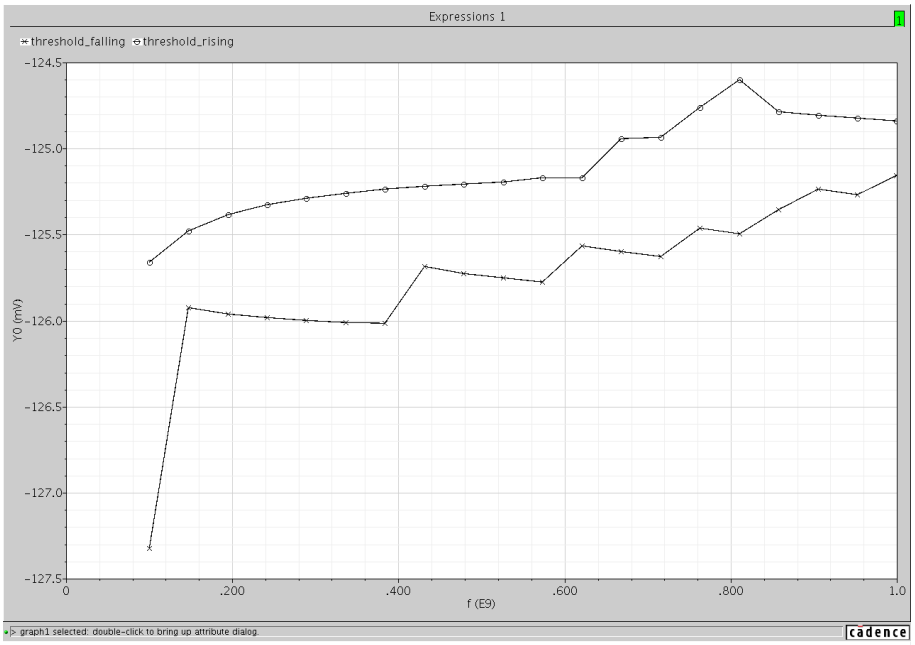


FIGURE 5.19: Simulated sampling frequency analyses.

It indicates that, with typical transistor's parameters, the sampling frequency can be increased. This is done mainly at expenses of area (wider transistors). The regeneration time in the latch is not critical because, as already mentioned, it was over-dimensioned, so no changes in this circuit were required.

It follows a *Monte-Carlo* analyses to account for effect of process and mismatch deviations of each transistor in the offset. For this, the UMC technology supplies statistical information for each parameter. In this simulation, the maximum deviation for process and mismatch parameters was set to 3 times the standard deviation (3σ is a typical value which covers 99.6% of the sampled values in a Gaussian distribution).

The result for 500 runs over a single C_{-125} comparator is presented in figure 5.20. The achieved mean value is $mu = \mu = -123.785mV$ and the standard deviation is $sd = \sigma = 4.09$. This indicates that 99.6% (3σ) of the simulated results are within the defined maximim offset of $15.625mV$.

Note that the mean value can be easily adjusted. The main result is the standard deviation, which is within the defined maximum. However, recalling that the required maximum

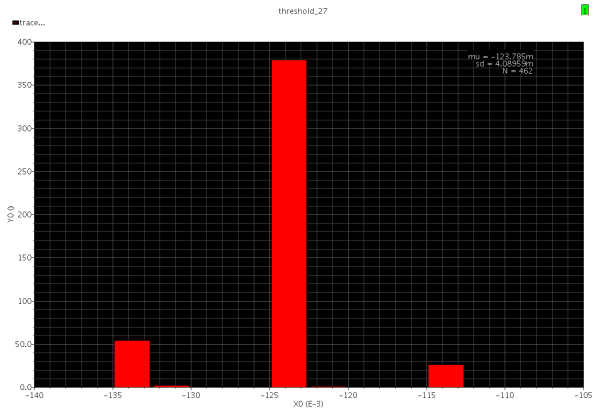


FIGURE 5.20: Monte-Carlo simulations for a C_{-125} comparator.

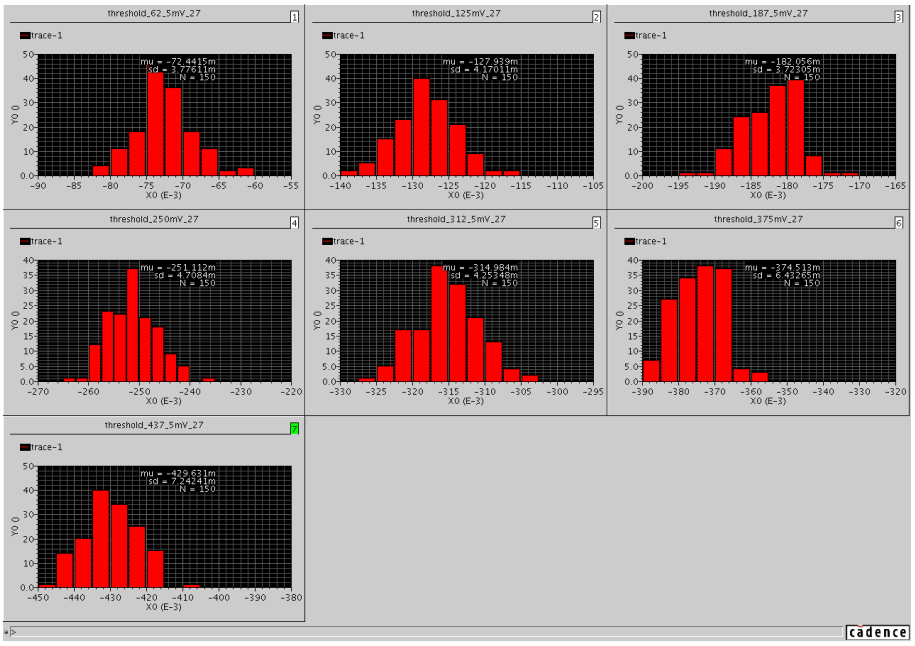


FIGURE 5.21: Monte-Carlo simulations for all different (7) comparators.

offset is $1/2^5 = 31.25mV$, and that the $15.625mV$ value was set to leave some headroom for unaccounted parasitics and mismatches, the achieved result is actually better than the initially expected.

This same conclusion can be made by analyzing the Monte-Carlo simulation for all the comparators in figure 5.21. However, for the higher comparison levels, C_{-375} and $C_{-437.5}$, the standard deviation increases. The reason for this is explained next.

In figure 5.22 (a) one can see that the variation of the total capacitance during the sampling

phase exhibits a flat region around $V_{DD}/2$, due to the complementary characteristics of the PMOS and NMOS transistors. The changes of the total capacitances that result in the parametric amplification during the hold/boost phase, can be observed in figure 5.22 (b), but only close to $V_{DD/2}$. In this phase, the capacitance reduction is significant, since both transistors entered into depletion, suggesting that the amplifier should operate in this region.

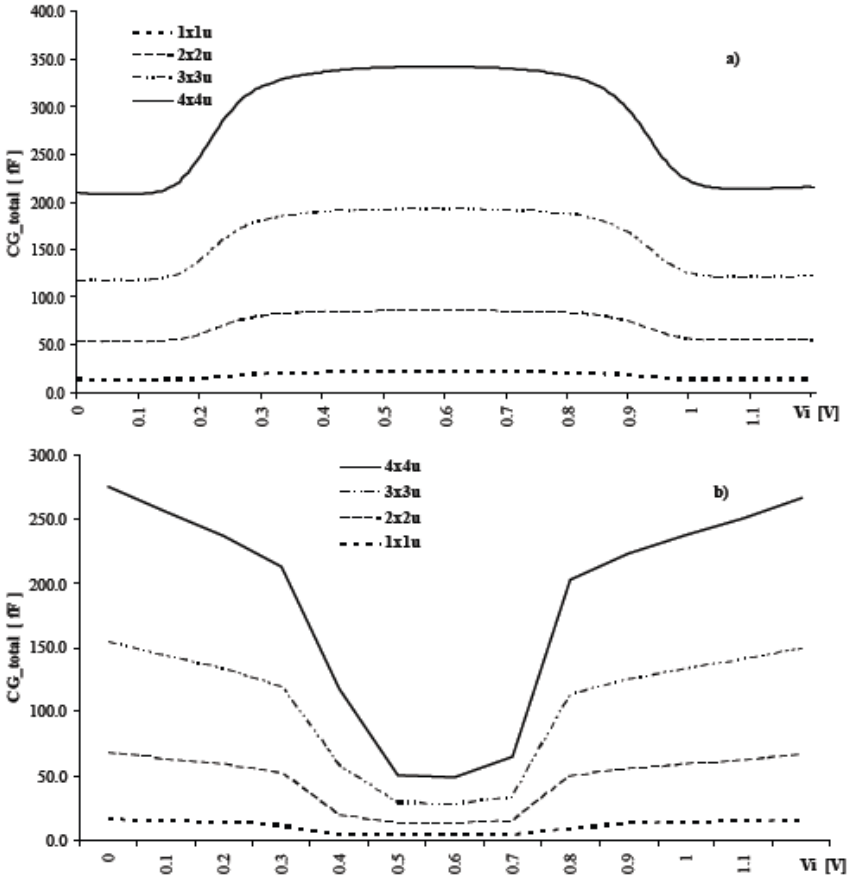


FIGURE 5.22: Simulated total gate capacitance in a MOScap as a function of V_{in}: a) sampling phase; b) hold/boost phase [22].

The C_{-375} and $C_{-437.5}$ operate too close to the acceptable limits, where the capacitance variance, from sampling to hold/boost phase, begins to decrease. As discussed in section 4.3.3, this results in less gain which increases the input offset. This effect should have some impact in the INL and DNL parameters and also in the harmonic distortion, as will be discussed next when analyzing the FFT.

The final simulation includes the complete proposed architecture and FFT analyses. The output of the complete circuit with a full scale input sine wave is presented in figure 5.23. The computed FFT for an input signal with 13 MHz is shown in figure 5.24.

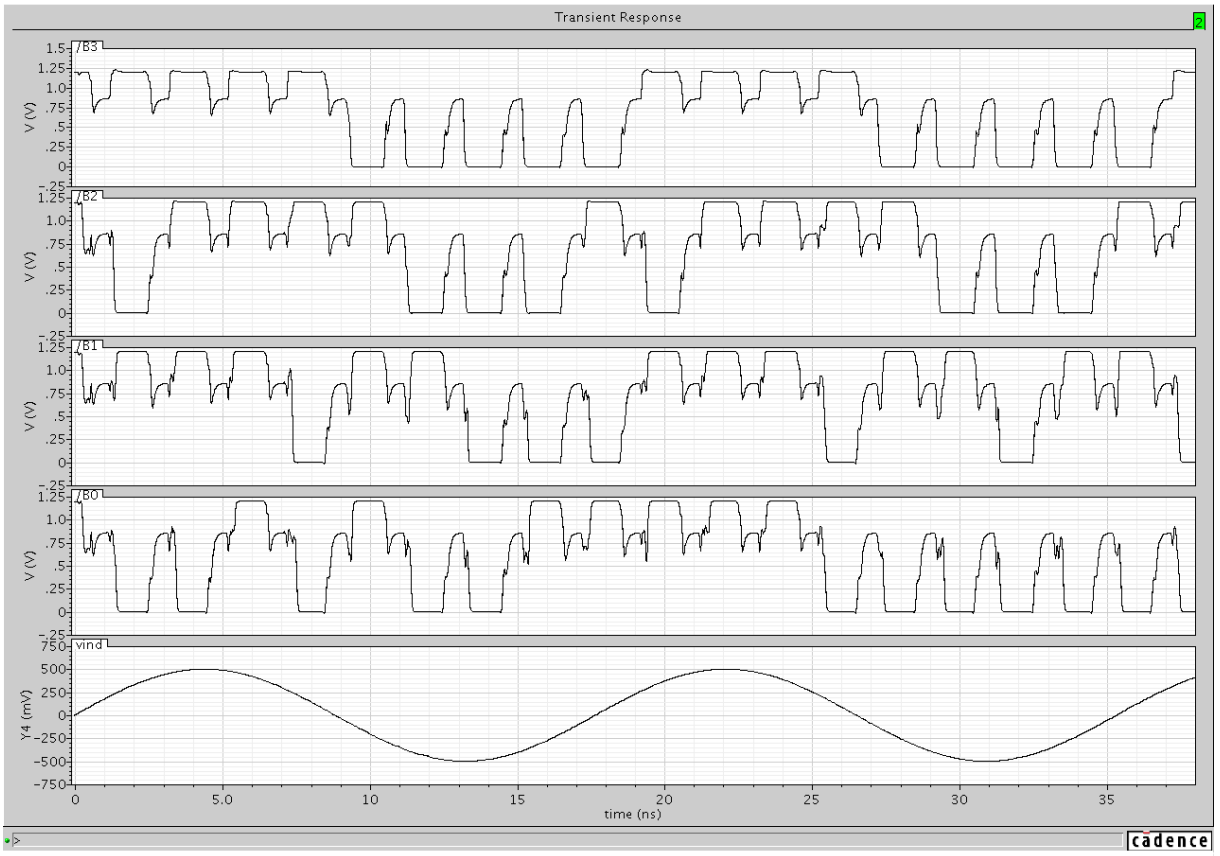


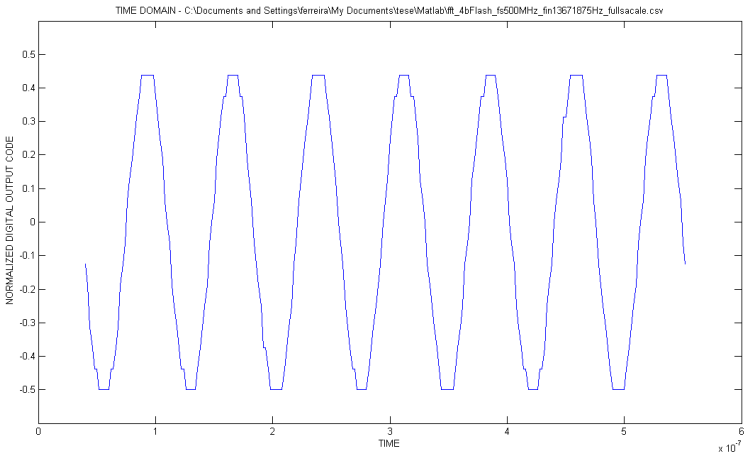
FIGURE 5.23: Simulated output of the complete Flash for an input sine wave.

Table 5.8 shows the achieved results for the dynamic parameters in the simulated ADC at different input frequencies and with $f_s = 500\,MHz$. For these, coherent sampling [24] have been used.

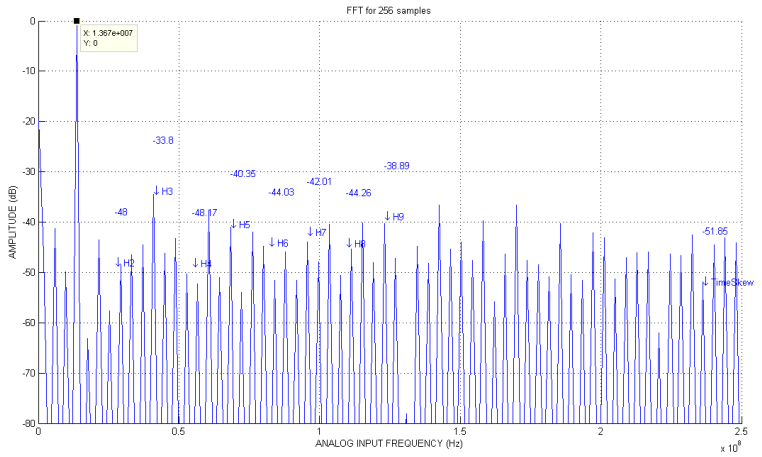
f_{in} @ 500 Ms/s	SNR_{dB}	THD_{dB}	$SFDR_{dB}$	$SNDR_{dB}$	ENOB
13.671875 MHz	27.422	-30.918	33.801	25.817	3.9962
56.640625 MHz	28.908	-28.630	34.199	25.757	3.9861
103.515625 MHz	28.118	-29.456	36.162	25.726	3.9810
154.296875 MHz	27.081	-31.814	35.868	25.882	3.9971
212.890625 MHz	28.013	-29.450	35.279	25.622	3.9701

TABLE 5.8: FFT dynamic specifications results

A qualitative analyses of this results indicates that the effective number of bits is limited by the quantization noise which, in a 4-bit ADC, defines a maximum achievable SNR of



(a) Reconstructed signal in time domain



(b) FFT

FIGURE 5.24: FFT analysis from simulated data (with no transient noise).

25.84 dB (B.11). However, notice that this is not a transient noise simulation and, hence, thermal noise is not included (KT/C and transistors noise).

The achieved SNR indicates that the quantization noise is limited to an interval with amplitude lower than 1 LSB, thus, increasing the number of bits to 5 should result in an ENOB higher than 4. However, one must remember that with transient noise the expected ENOB will be lower than the one achieved without it.

This result has to be confirmed because the THD may be limiting the ENOB if the

quantization noise decreases (by the additional bit). Note that figure 5.24(a) shows sharp corners when the sine wave reaches maximum and minimum amplitude. This results in harmonic distortion and can be associated with the C_{-375} and $C_{-437.5}$ comparators.

By analyzing the results from the Monte-Carlo analyses in figure 5.21, the maximum INL and DNL parameters can be found. Accounting for $3\,\sigma$ yield or 99.6 % of the cases and assuming that the threshold can be adjusted in future work, one can achieve:

$$INL_{max} = \frac{3 \times 7.34}{62.5} \approx 1/3\,LSB \tag{5.3}$$

$$DNL_{max} = \frac{3 \times 7.34 + 3 \times 6.43}{62.5} \approx 1.5\,LSB \tag{5.4}$$

This results are not good at all. They both should be below $1/2\,LSB$. However, recalling again the problems associated with the C_{-375} and $C_{-437.5}$ comparators, one can analyze the INL and DNL if this were corrected. If so, the achieved results would be:

$$INL_{max} = \frac{3 \times 4.7}{62.5} \approx 1/4\,LSB \tag{5.5}$$

$$DNL_{max} = \frac{3 \times 4.7 + 3 \times 4.2}{62.5} \approx 1/2\,LSB \tag{5.6}$$

This last result indicates that a 5 bit ADC can be achieved if the problems associated with the C_{-375} and $C_{-437.5}$ comparators can be minimized.

5.5 1.5 and 2 bits parametric Flash Quantizer for pipeline ADC stages

The 8 bit, $120MHz$ pipeline ADC presented in [1] required 1.5 bits and 2 bits flash converters. The previous study described in this work was adapted in order to provide this ADCs. The low sampling frequency means that the input switches can be made smaller, since the time constant can be bigger, and the dynamic latch regeneration time

can be extended. The lower number of bits reduces the required accuracy and allow smaller sampling capacitances and smaller transistors in the dynamic latch.

The layout shown in figure 5.25 shows the complete die and pads (top-left figure), with a total area of $2.25mm^2$, the complete pipeline (top-right figure) with $0.12mm^2$ and three side-by-side comparators to implement the 2 bit flash converter. Each comparator has an area of $660\mu m^2$.

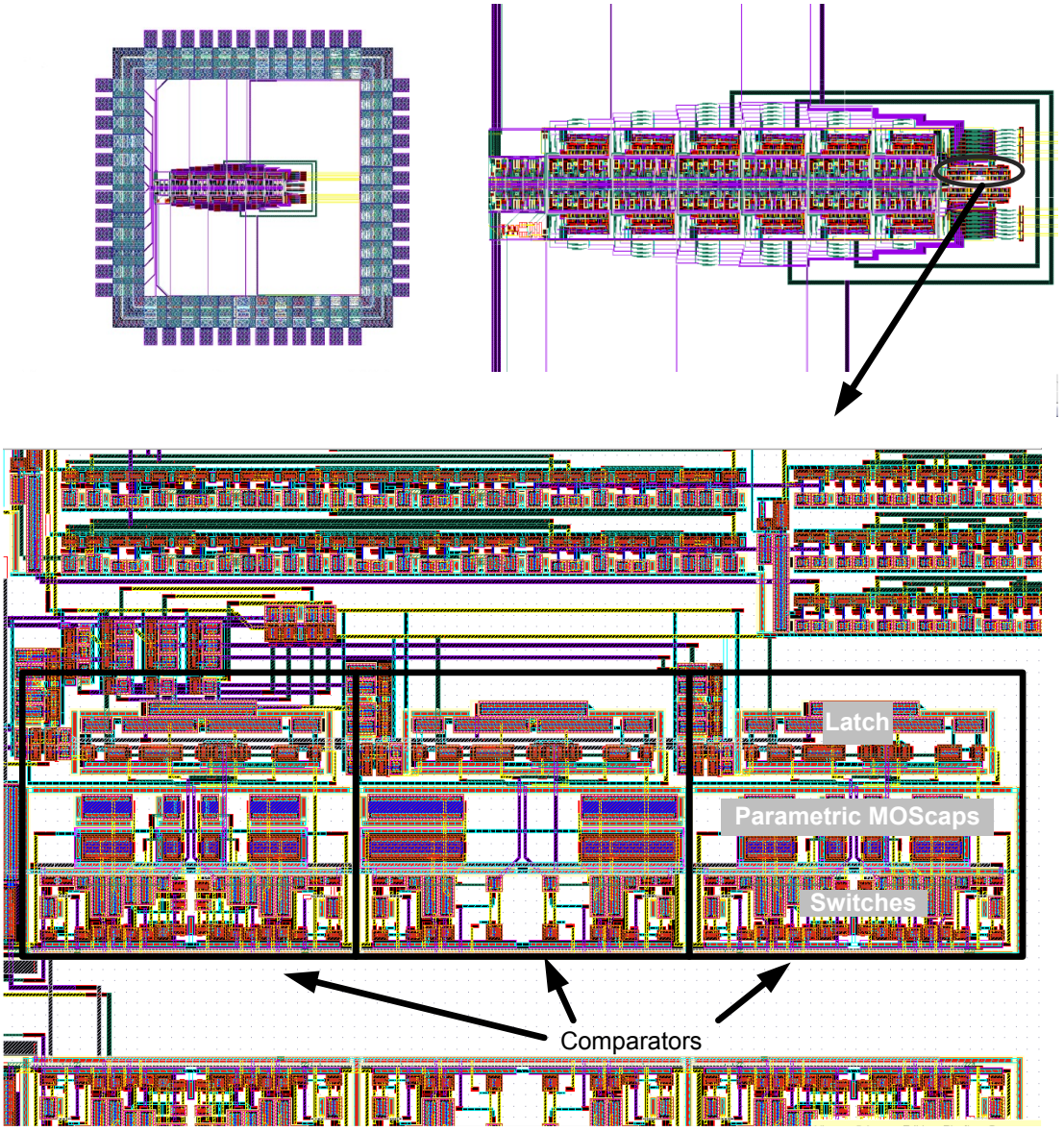


FIGURE 5.25: Comparators in an 2 bit pipelined architecture (layout).

5.6 Conclusions

This brief has presented a MOS only 4-bit 500-MS/s Flash ADC that extensively uses Parametric pre-amplification. Measurements for a input signals within the $f_s/2$ range indicate that the effective number of bits is constant and equal to 4, indicating that a 5 bit flash is possible. In the static parameters, the identification of two troublesome comparators indicates that the INL and DNL parameters are within the requirements.

In order to have some silicon results, two quantizers, respectively with 1.5 bit (2 comparators) and with 2 bits (3 comparators) were design (the layout is presented in figure 5.25). These were used in the 7 pipelined stages of an 8 bit time-interleave pipeline ADC with digital correction.

Measured results of 3 prototypes demonstrate the functionality of the proposed parametric based comparator's architecture. Their results are out of the scope of this work and can be found in [1].

Chapter 6

Conclusions and Future Research

The complete receiver is the obvious future work. For this an Automatic Gain Control between the LNA and Mixer, and between the Mixer and ADC should be included. Moreover, filters need some extensive attention.

The use of the Flash architecture however allows, with minor adjustments, to turn on or off some comparators to achieve better power efficiency in systems where the 4 bits are not necessary all the time. For example, Ultra-Wide Band impulse radio (UWB-IR) systems uses extremely short pulses with duration of the order of nanoseconds to transmit information. In practice, the receiver needs only to detect this pulse, and for this a single comparator (1 bit) can preform the task if the signal is strong. For this type of signals, with very large bandwidth, the sampling frequency must be in the order of GHz, so Flash converters are, most likely, the best choice.

The parametric amplification used in the comparator should be tested. Measurements with and without this feature are required to evaluate the actual performance of this amplification. By recalling the problems observed in the C_{375} and $C_{437.5}$ in the simulations preformed, the indications are that it works, since these two suffer from less gain in the parametric amplification.

Moreover, the benefits of parametric amplification can be applied to other circuits. Passive mixers have lower (and no static) power dissipation and a much smaller flicker noise than active mixers (because there is no DC current). In CMOS technology the flicker noise limits the circuit performance in DCR, thus Low-IF was the choice in this work.

The use of passive mixers can minimize this problem, however, these don't have conversion gain, which affects the overall gain and noise figure. To overcome this limitation, a passive SC mixer with parametric amplification is already being analyzed and the work entitled "MOSFET-Only MIXER/IIR Filter With Gain using Parametric Amplification" has been accepted for ISCAS'10.

The simulated results in the ADC indicate that this architecture can achieve a minimum of 5-bits. Moreover, the over-dimensioned latch transistors can be tuned for a better power efficiency. Finally, the simulations at 1GHz indicate the possibility of higher sampling rate, making it more adequate for UWB applications.

The mixer's simulation results can be improved with the use of a differential inputs. An on-going work is an optimized balun-LNA to be used in this architecture.

All of this points demand for a deeper study and validation. In fact, one of the most important issues that is yet unknown is the actual removal of charges during the parametric amplification. The time required for this processes may not be properly accounted by the simulation. Thus, to achieve higher sampling speed, implementation and measurements are required.

This previous considerations leave an open door for multiple future analysis.

Appendix A

Important Measurements in Mixers

A.1 CG

The CG (Conversion Gain) in mixers is defined as the ratio between the power delivered to the load (at IF) and the power delivered by the source (in the RF).

A.2 NF

The noise figure is a measurement of the noise added by the circuit, expressed in dB. It is defined as the ration of the total noise at the output and the noise at the input multiplied by the conversion gain.

A.3 1dB Compression point

This is a measure of the circuit linearity. It is defined as the output signal power that corresponds to a difference of 1 dB from the ideal (linear) circuit.

A.4 IIP3

The IIP3 (third-order Input Intercept Point) relates nonlinear products caused by the third-order nonlinear term to the linearly amplified signal.

The intercept point is obtained graphically by plotting the output power versus the input power both on logarithmic scales. Two curves are drawn one for the linearly amplified signal at an input tone frequency and another for a nonlinear product.

Both curves are extended with straight lines. The point where the curves intersect is the intercept point. It can be read off from the input or output power axis, leading to input or output intercept point, respectively (IIP3/OIP3).

Input and output intercept point differ by the small-signal gain of the device.

Appendix B

Important Measurements in ADCs

B.1 Static Specifications

B.1.1 INL and DNL

The INL (integral non-linearity) is the difference between the ideal code transition point and the actual one. It usually defined in LSB units.

The INL is calculated for each transition point k, and the maximum value achieved is the ADC INL.

The DNL (Differential non-linearity) is the difference between the the ideal step width and the actual one. It is also define in LSB units.

The DNL is calculated for each step k, and the maximum value achieved is the ADC DNL.

ADCs with large INL show harmonic distortion. Large DNL lead to INL with large random components. The error is equivalent to noise added in the quantization noise that degrades the SNR [15].

B.1.2 Monotonicity

An ADC that produces output codes that are consistently increasing with the increase of the input and consistently decreasing with the decrease of the input is said to be monotonic.

B.1.3 Missing Code

A missing code is a code that an ADC never produces. This can occur in ADCs with large DNL.

B.2 Dynamic Specifications

B.2.1 SNR

The Signal-to-Noise Ratio (SNR) is a measure that quantifies the relationship between the wanted signal's power and the total noise that corrupts that signal, within the desired bandwidth.

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{B.1}$$

Where P is the average power. The most common form to express SNR is to define it in terms of the logarithmic decibel (dB) scale:

$$SNR_{dB} = 10\, log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{B.2}$$

Assuming that both signal and noise can be measured across the same impedance, one can achieve the SNR using:

$$SNR = \left(\frac{A_{signal}}{A_{noise}}\right)^2 \tag{B.3}$$

or, in dB:

$$SNR = 20\,log_{10}\left(\frac{A_{signal}}{A_{noise}}\right) \tag{B.4}$$

Where A is the *root-mean-square* amplitude.

B.2.2 SINAD or SNDR

The SNDR (Signal to Noise and Distortion Ratio) is a similar in definition to the SNR except that the non-linear distortion terms, generated by the input sine wave in a non-linear system, are also accounted.

B.2.3 ENOB

ENOB (Effective Number Of Bits) measure the same as the SNDR, but in bits. For a sine wave input these two are linked by:

$$ENOB = \frac{SNDR_{dB} - 1.76}{6.02} \tag{B.5}$$

B.2.4 HD

HD (Harmonic Distortion) is the ratio between the root-mean-square (rms) of the signal and the rms of harmonic components, including aliased terms.

B.2.5 THD

The THD (Total Harmonic Distortion) is the ratio between the rms value of the fundamental signal and the mean value of the root-sum-square of its harmonics.

B.2.6 SFDR

SFDR (spurious Free Dynamic Range) is the ratio between the rms value of the input signal amplitude and rms of the highest spurious spectral component in the first Nyquist zone. It is similar to the THD but focus only in the worst tone.

This is specially important in communication systems because high spurs from adjacent channels may fall very close to the wanted signal and mask it.

B.3 Quantization Noise

Quantization is an unavoidable source of error when converting an analog signal, continuous in time and amplitude, to a digital code, discrete both in time and amplitude. Since there is a finite number of bits (codes) to represent a given amplitude, an ADC defines an *interval* of input amplitudes for each output code.

Assuming $V_{FS} = V_{max} - V_{min}$, the amplitude of each interval, or *Quantization Step*, Δ is:

$$\Delta = \frac{V_{FS}}{M} \tag{B.6}$$

Where M is the number of quantization intervals.

Particularly useful while measuring quantization error, a new unit[1] is defined:

$$LSB = \frac{1}{2^n} \tag{B.7}$$

Where n is the number of bits and LSB stands for *least significant bit*.Note that, a change in the LSB corresponds to a voltage change of $V_{LSB} = \frac{V_{ref}}{2^n}$.

Figure B.1 illustrates the inputs and outputs of an ADC.

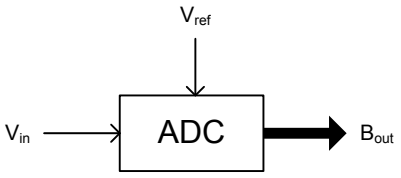


FIGURE B.1: ADC's inputs and output

These signals are related by:

[1]This unit is, in fact, unitless.

$$V_{in} = V_{ref}\left(b_1 2^{-1} + b_2 2^{-2} + \ldots + b_n 2^{-n}\right) \pm V_{\epsilon Q} \tag{B.8}$$

$$= V_{ref} B_{out} \pm V_{\epsilon Q} \tag{B.9}$$

Note that equation B.9 shows a relation between a digital code and a voltage, $V_{ref}B_{out}$. This is, in fact, the Digital-to-Analog converter's (DAC) work. This is important to properly analyze Quantization Error in ADCs.

The quantization error, $V_{\epsilon Q}$, accounts for the difference between $V_{ref}B_{out}$ and V_{in}, as figure B.2 illustrates, and goes to zero only when the number of bits goes to infinity.

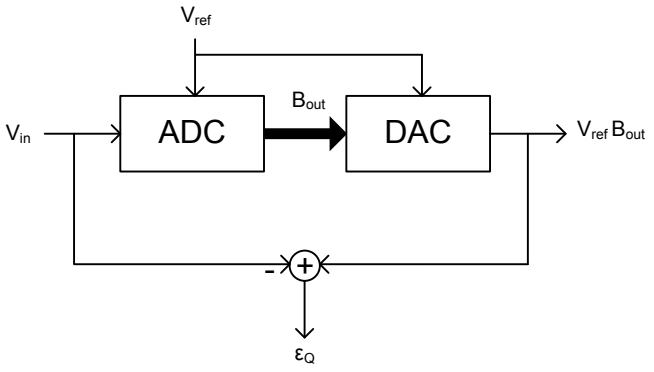


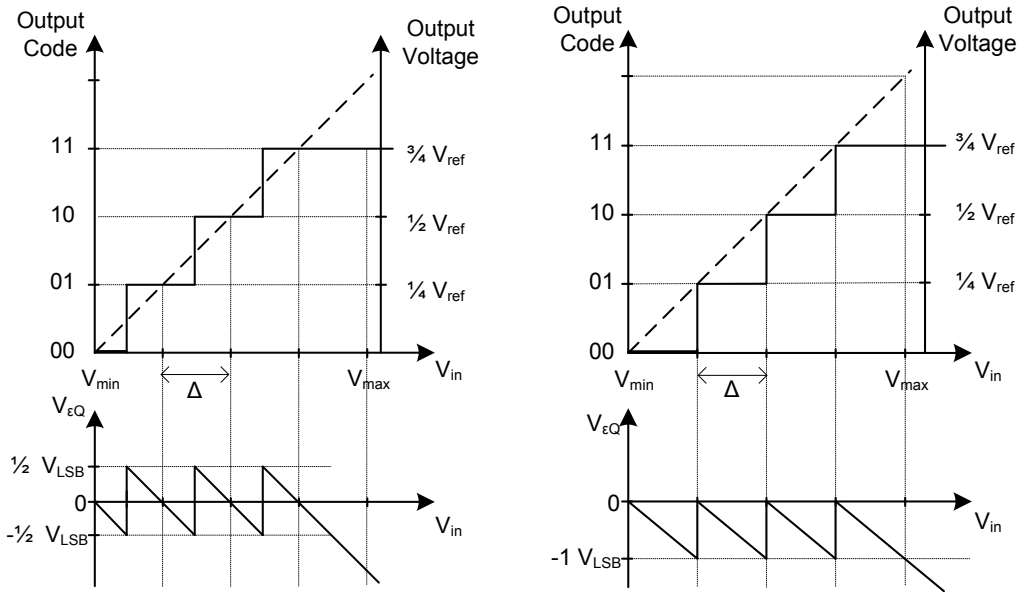
FIGURE B.2: Quantization Error

A common procedure to analyze the behavior of $V_{\epsilon Q}$ is to assume that the input voltage V_{in} is a ramp. In figure B.3(a), the quantization interval is defined by the *mid-point* of the quantization step (the output changes at the middle of Δ).

Figure B.3(b) is an equivalent representation but defines the quantization interval by it's lower edge.

Note that, if $V_{min} < V_{in} < V_{max}$ the quantization error is limited to a dynamic range of 1 LSB in both cases. This is still true for different input signals. Outside of the input dynamic range the output saturates and the quantization error increases.

By assuming that the input signal varies rapidly between V_{max} and V_{min}, one can consider that $V_{\epsilon Q}$ is a random variable, uniformly distributed between $\pm V_{LSB}/2$. The probability density function for such signal is a constant value.



(a) Quantization Error when the quantization interval is defined by the mid-point

(b) Quantization Error when the quantization interval is defined by the lower edge

FIGURE B.3: Quantization error when V_{in} is a ramp.

As a result, the rms value of quantization error is given by [14]:

$$V_{\epsilon Q(rms)} = \frac{V_{LSB}}{\sqrt{12}}. \tag{B.10}$$

Assuming a sinusoidal input signal at the input with $V_{in(rms)} = \frac{\sqrt{2}}{2}(V_{max} - V_{min})$, results in a usefull expression that indicates the best possible SNR that an n-bit ADC can achieve, i.e., if no other noise source degrades the signal (which is impossible), the SNR is limited by the quantization error.

$$SNR = 20log\left(\frac{V_{in(rms)}}{V_{\epsilon Q(rms)}}\right) = 6.02n + 1.76 \tag{B.11}$$

B.4 kT/C Noise

Another limit for of any data converter system is the kT/C noise. It occurs during the sampling procedure and is associated to the finite conductance of input switch, which

results in thermal noise added to the sampling capacitance.

During the sample phase, the RC network equivalent noise model can be represented by the circuit in figure B.4.

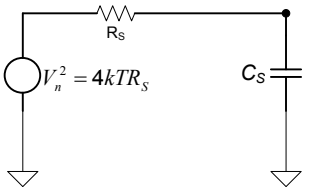


FIGURE B.4: RC network noise model

Where k is the *Boltzmann constant* , T is the absolute temperature (in Kelvin) and R_S account for the resistance of the input voltage source in series with the input switch's on-resistance.

When the input switch goes off, the sampling capacitance hold not only the input voltage but also the noise. The spectrum of v_{n,C_s} is given by the $4kTR_s$ spectrum filtered by the R_sC_s network transfer function.

$$v_{n,C_s}(\omega) = \frac{4kTR_s}{1 + (\omega R_s C_s)^2} \tag{B.12}$$

Therefore, the total noise stored on C_s during the sample phase is:

$$P_{n,C_s} = \int_0^\infty \frac{4kTR_s}{1 + (\omega R_s C_s)^2} = \frac{kT}{C_s} \tag{B.13}$$

Note that the total noise added by the resistance is independent of the resistance value. This is because the added noise caused by increasing the resistance is balanced by the lower cut-off frequency of the RC filter.

Since absolute zero temperature or infinite capacitance can not be achieved, to keep this noise source negligible, the sampling capacitance must be made big, which translates in larger die area.

The sampling operation is inherent to all ADC circuits, one can realize that this noise source will add to the quantization noise and reduce the best possible SNR of equation B.11.

Appendix C

Published Papers

C.1 TCAS-II, Feb.2010

An 8-bit 120-MS/s Interleaved CMOS Pipeline ADC Based on MOS Parametric Amplification

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS

An 8-bit 120-MS/s Interleaved CMOS Pipeline ADC Based on MOS Parametric Amplification

J. Oliveira, *Member, IEEE*, J. Goes, *Senior Member, IEEE*, M. Figueiredo, *Student Member, IEEE*, E. Santin, *Student Member, IEEE*, J. Fernandes, *Member, IEEE*, and J. Ferreira

***Abstract*—This brief presents an 8-bit 120-MS/s time-interleaved pipeline analog-to-digital converter (ADC) fully based on MOS discrete-time parametric amplification. The ADC, fabricated in a 130-nm CMOS logic process, features an active area below 0.12 mm^2, where only MOS devices are used. Measurement results for a 20-MHz input signal shows that the ADC achieves 39.7 dB of signal-to-noise ratio, 49.3 dB of spurious-free dynamic range, −47.5 dB of total harmonic distortion, 39.1 dB of signal-to-noise-plus-distortion ratio, and 6.2 bits of peak effective number of bits while consuming less than 14 mW from a 1.2-V supply.**

***Index Terms*—Analog-to-digital converter (ADC), MOS parametric amplification, pipeline, time-interleaved.**

I. Introduction

TECHNOLOGY scaling is raising many issues for analog circuit design, such as intrinsic device gain, supply voltage, and device variability. In particular, for switched-capacitor (SC) realizations, performance degradation increases the difficulty to realize accurate charge transfers in the traditional manner, i.e., based on closed-loop operational amplifier (opamp) structures. Additionally, to reach a higher operation frequency, these opamps need a high gain–bandwidth product, reflecting an increase on power consumption. Different alternative solutions have recently been proposed, namely, open-loop amplification [1], comparator based [2], zero-crossing based [3], MOS discrete-time (DT) parametric amplification (MPA) [4], and dynamic source follower (SF) based [5]. With the exception of MPA, all of these techniques have already been applied to pipeline analog-to-digital converters (ADCs). However, their energy and area efficiency values are much dependent on the optimum resolution per stage, scaling of capacitance values, and residue amplifier topology for each multiplying digital-to-analog converter (MDAC) block.

Manuscript received July 9, 2009; revised October 6, 2009 and November 7, 2009. This work was supported in part by the Portuguese Foundation for Science and Technology (FCT/MCTES) under LEADER (PTDC/EEA-ELC/69791/2006) and SPEED (PTDC/EEA-ELC/66857/2006) Projects. This paper was recommended by Associate Editor H.-S. Chen.

J. Oliveira, M. Figueiredo, E. Santin, and J. Ferreira are with the Centre of Technology and Systems, Institute for the Development of New Technologies (CTS-UNINOVA), and the Department of Electrical Engineering, Faculty of Sciences and Technology (FCT), Universidade Nova de Lisboa (UNL), 2825-114 Monte da Caparica, Portugal (e-mail: jpao@fct.unl.pt).

J. Goes is with the Department of Electrical Engineering, FCT, UNL, 2825-114 Monte da Caparica, Portugal, and also with Silicon and Software Systems (S3), Campus FCT/UNL, 2829-516 Caparica, Portugal.

J. Fernandes is with the R&D IC Unit, INESC-ID, 1000-029 Lisbon, Portugal (e-mail: jorge-fernandes@inesc-id.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSII.2009.2038632

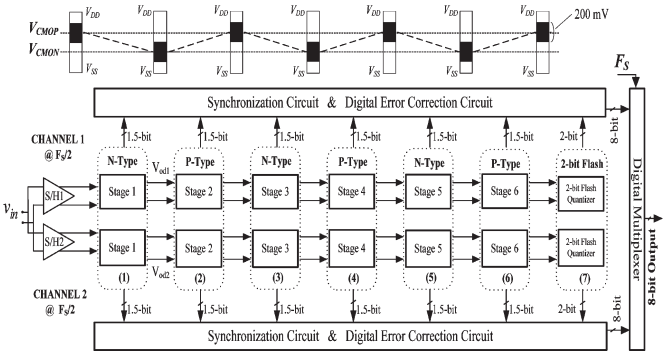


Fig. 1. ADC architecture.

Silicon realizations using the DT MPA concept have already been demonstrated in a low-speed SC amplifier [4] and in a finite-impulse response low-pass filter circuit internally operating at a clock rate of 80 MS/s [6]. Notice also that the technique proposed in [5] does not use any kind of MPA.

This brief describes a complete 8-bit time-interleaved pipeline ADC that uses the concept of MPA in all the analog blocks of the architecture. The simplicity of the basic and original MPA structure [4] is preserved, but to reach higher gains, one terminal of the MOS capacitor (MOSCAP) is left floating. In addition, in contrast to other solutions that, in some way, include metal capacitor structures in their design, this ADC only uses MOS devices, and therefore, it is fully compatible with standard CMOS digital processes. Moreover, based on the authors' knowledge, this work describes the first silicon proven high-speed pipeline ADC using the DT MPA concept.

This brief is organized as follows. In Section II, the adopted ADC architecture is presented. In Section III, the DT MPA principle is revisited, and a circuit modification is proposed. In Section IV, each building block is presented, explicitly showing where the MPA technique has efficiently been applied. In Section V, a simplified ADC noise analysis model is described. Section VI shows the obtained experimental results, and finally, in Section VII, the conclusions are gathered.

II. ADC Architecture

The two-channel interleaved ADC is shown in Fig. 1. Each channel operates at half of the conversion rate, i.e., $F_S/2$, and it comprises a sample-and-hold (S/H) followed by six pipelined stages with minimum resolution (1.5 bits) and by a 2-bit Flash quantizer (FQ). The 14 output bits are digitally synchronized, and 8 bits are available after digital correction. Pipeline stages consist of two types, namely, N and P. These two versions are needed to efficiently deal with different input and output common-mode (CM) voltages resulting from each

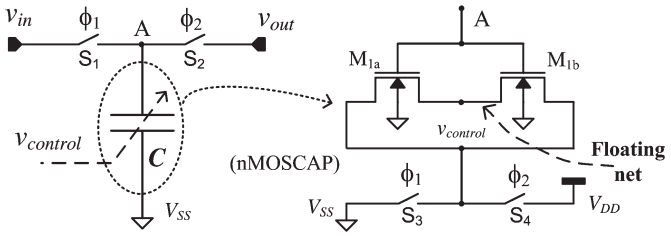


Fig. 2. MOS parametric amplifier using two separated devices.

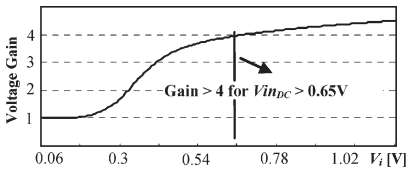


Fig. 3. Parametric amplifier gain as a function of the input dc level.

type of stage, which affect the gain in the MPA structures. For example, N-type stages sample an input signal with a higher input CM voltage and produce a signal with a lower output CM voltage, i.e., $V_{\rm CMO}$. P-type stages work in the opposite way. The stages process differential signals up to 400 mVpp.

III. Modified MOS Parametric Amplifier

Small values of signal voltage amplification, e.g., to implement a multiply-by-two residue amplifier (MBTA), can be achieved by using the parametric MOS structure described in [4], where a DT amplifier is evaluated rather than a continuous-time configuration, as analyzed in [7]. In this amplifier, the gain is set through the reduction of the total equivalent gate capacitance of a single MOSCAP device while maintaining the total gate charge between the sampling phase ϕ_1 and the amplification phase ϕ_2. As explained in [4], the capacitance reduction of a MOSCAP can be achieved by moving it from inversion into depletion, as result of changing the control voltage $V_{\rm control}$ applied to the drain from the negative power supply voltage $V_{\rm SS}$ to the positive power supply $V_{\rm DD}$, as shown in Fig. 2.

The first difference of the MOSCAP structure proposed in this work from the original structure used in [4] lies on the fact that two half-sized MOSCAPs are used in parallel rather than a single MOSCAP and with one of the tied terminals left floating. The division into more than two devices can be used for higher unit capacitance values. Care must be taken in not using either large L (due to speed limitations) or minimum L (due to short-length effects). As a consequence, with this modified structure, it becomes possible to decrease the effect of the extrinsic (overlap) gate capacitances during the amplification phase, i.e., ϕ_2. Hence, amplification gains above 2 (considering the loading effect) can now be easier achieved with an nMOS-type MOSCAP (nMOSCAP). By properly adjusting the load, it becomes possible to design MBTA circuits with gain accuracy values above the 6-bit level, without a calibration scheme.

The amplifier gain shown in Fig. 3 was determined using the intrinsic MOS gate capacitance values obtained from electrical simulations of the circuit shown in Fig. 2. Fig. 3 also shows that the maximum achievable gain also depends on the CM level of the input voltage, reflecting how well the MOS device is biased in the inversion region during the sampling phase. Therefore, an appropriate dc level has to be carefully chosen.

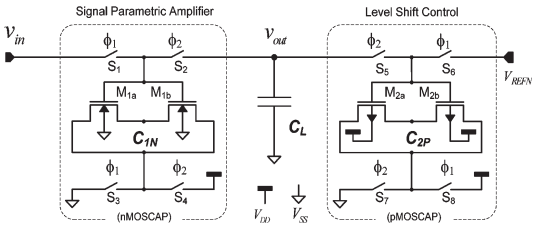


Fig. 4. CMOS DT amplifier with output level-shift control.

Another problem to be solved is related with the existing dc level shifting that occurs when an nMOSCAP, i.e., C_{1N}, changes from inversion into depletion. To avoid this, since during the amplification phase this level tends to rise above $V_{\rm DD}$, an additional pMOS-type MOSCAP (pMOSCAP), i.e., C_{2P}, is added, as shown in Fig. 4, to produce the opposite effect (negative dc level shifting).

A first-order charge redistribution analysis applied to the described amplifier results in

$$v_{\rm out} \approx \frac{\alpha_{C_{1N}}}{k} v_{\rm in} + \frac{\alpha_2 - \alpha_4}{k} V_{\rm DD} + \frac{\alpha_4}{k} V_{\rm REFN} + \frac{\alpha_3}{k} V_{C_L,\phi_1} \quad (1)$$

where $C_{1N,\phi_1} = \alpha_{C_{1N}} \cdot C_{1N,\phi_2}$, $C_{2P,\phi_2} = \alpha_2 \cdot C_{1N,\phi_2}$, $C_L = \alpha_3 \cdot C_{1N,\phi_2}$, $C_{2P,\phi_1} = \alpha_4 \cdot C_{1N,\phi_2}$, and $k = 1 + \alpha_2 + \alpha_3$. k models the gain reduction due to the load capacitance C_L and C_{2P} during phase ϕ_2. α_{C1N} reflects the capacitance variation from phase ϕ_1 to phase ϕ_2 for C_{1N}. α_2 represents the relation between C_{2P} and C_{1N} in phase ϕ_2. The remaining parameters α_3 and α_4 describe the ratio of C_L and C_{2P} in phase ϕ_1 with respect to the value of C_{1N} during phase ϕ_2.

Equation (1) demonstrates that, in addition to the desired multiplying factor, the output voltage has an offset component partially controlled by the pMOSCAP. It also shows that the load capacitance C_L affects the attained circuit gain and the output offset level (depending on how the signal is applied to C_L during ϕ_1, V_{C_L,ϕ_1}). To reduce this loading capacitance effect, a simple SF can be inserted at the output node, but if not accurately sized, it could be a source of unnecessary power dissipation. This SF also buffers the output when loaded.

IV. MPA-Based ADC Design

Each pipelined stage uses the MPA technique, as shown in Fig. 4, in the 1.5-bit MDAC (following an open-loop approach) and in the comparators employed in the local 1.5-bit FQ.

A. Pipeline Stage of 1.5 bit

An N-type stage, as depicted in Fig. 5, has vertical symmetry, due to the interleaved operation. Each channel comprises a 1.5-bit FQ, a 1.5-bit MDAC composed of two half-MDACs (HMDACs) blocks, and two SFs. Both channels share the same replica bias circuit (RBC) to adjust the CM voltage $V_{\rm CMO}$ of the stage, thus avoiding accumulation of dc errors through the pipeline chain and minimizing the impact of PVT variations. A similar configuration is used for the P-type stage.

B. RBC for Output CM Control

The RBC, as shown in Fig. 6, consists of four interleaved downscaled (by 2) HMDACs, for power and area savings, connected to a single (replica) SF circuit (downscaled by 2).

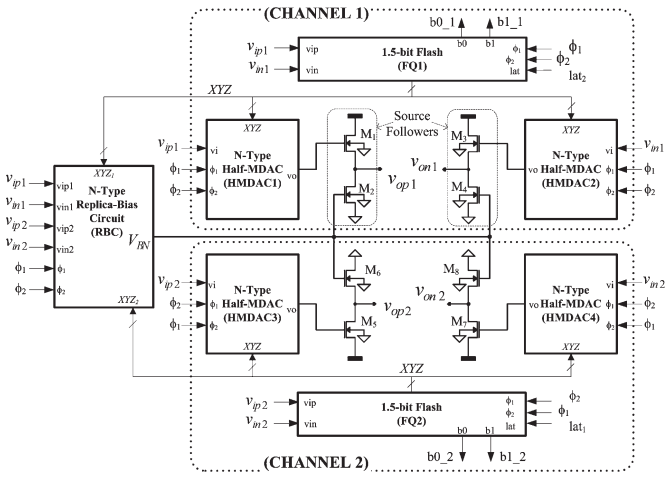


Fig. 5. Block diagram of one interleaved pipelined stage (N-type).

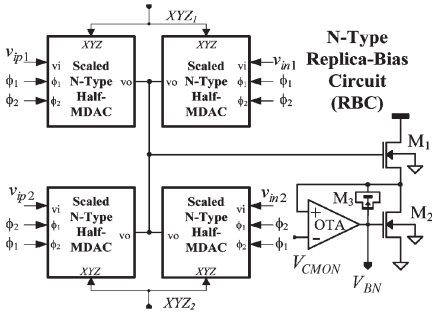


Fig. 6. RBC (N-type).

The averaged voltage at the SF's output is set and controlled by a feedback loop using a low-gain low-power operational transconductance amplifier (OTA; < 30 μW). The RBC provides the required regulated biasing voltage $V_{\rm BN}$ to the main SFs used in the output of each stage. The black filled areas of the voltage bars in Fig. 1 represent the allowed single-ended voltage levels at the different nodes, and the dashed line represents the $V_{\rm CMO}$ variations between stages. The $V_{\rm CMO}$ of an N-type stage is set to a lower value ($V_{\rm CMON} \approx 0.35$ V) to properly drive the next P-type stage. Likewise, the $V_{\rm CMO}$ of a P-type stage is set to a higher value ($V_{\rm CMOP} \approx 0.75$ V) to drive the next N-type stage. The OTA does not need a high bandwidth since it is not inserted in the main signal path. It can be designed with minimum current.

C. HMDAC Circuit

The schematic of an N-type HMDAC circuit [8] for the positive signal path is shown in Fig. 7. For the P-type version, the nMOSCAP and the pMOSCAP are interchanged. Moreover, the only difference between the positive and negative signals (either in the N- or P-type stage) is that signals X and Z from the local FQ are exchanged between the two HMDACs. MOSCAPs C_X, C_Y, and C_Z, operating in the inversion region, perform the DAC function. C_Y is required to provide the same loading, although it does not add any charge (differentially).

Only one of these capacitors is selected in each clock cycle for charge redistribution by the X, Y, and Z codes provided by the local FQ. Due to charge conservation, the input signal applied to the HMDAC block is amplified by the

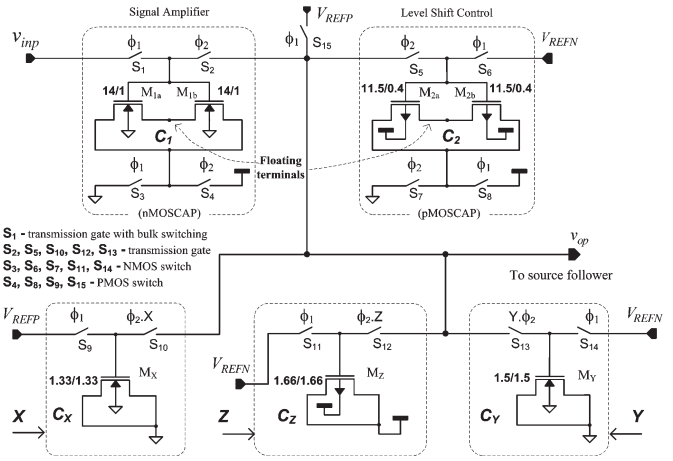


Fig. 7. N-type HMDAC. W/L values are given in micrometers/micrometers.

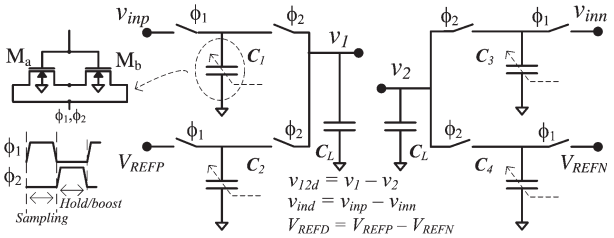


Fig. 8. Input sampling network with embedded amplification.

MPA principle through C_1. This process adds a positive dc level shifting. As previously stated, to prevent this shift from exceeding $V_{\rm DD}$, a MOSCAP C_2 (pMOS) is used to control the dc level. By optimizing the size of all five MOSCAPs (sizes in Fig. 7), the MDAC function is implemented at the SFs' outputs, i.e., $v_{\rm outd} = 2 \cdot v_{\rm ind} - X \cdot V_{\rm REFD} + Z \cdot V_{\rm REFD}$. During ϕ_1, the sampling capacitor C_1 has a capacitance value of 312 fF (in inversion). This value roughly drops by a factor of 5 during ϕ_2 when C_1 goes into depletion. Reference voltages $V_{\rm REFP} = 1$ V and $V_{\rm REFN} = 0$ V are adopted to minimize C_X, C_Y, C_Z, and C_2. The MPA cell has been designed to tolerate $\pm 5\%$ $V_{\rm DD}$ variations and still keeping the residue amplification gain with an error below $\pm 2.5\%$ (for over 6-bit input-referred accuracy). Finally, the front-end S/H is simply implemented by a P-type MDAC with its variable Y enabled.

D. FQs of 1.5 bits

Each 1.5-bit FQ comprises two comparators, two digital latches, and an output encoder to provide the X, Y, and Z signals and the output bits. Each comparator consists of an input SC network that defines the desired threshold levels of $\pm V_{\rm REFD}/4$, followed by a positive-feedback latch (PFDL). As proposed in [9], to improve performance and reduce power, preamplification is embedded in the input SC network also by employing the MPA principle, as shown in Fig. 8. The circuit operates with two nonoverlapping phases, and each capacitor is implemented by means of a parametric MOSCAP, as described in Fig. 2. A detailed analysis is given in [9].

The threshold level does not directly depend on the load capacitance and can be adjusted by the ratio of C_1 and C_2 during ϕ_1. The load capacitance C_L only affects the circuit gain. The complete comparator schematic is shown in Fig. 9.

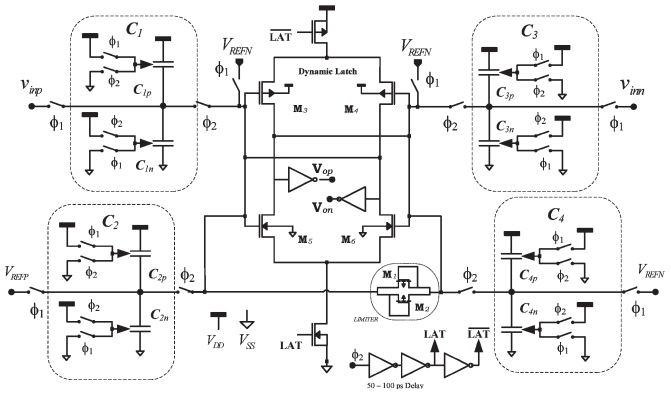


Fig. 9. Comparator with MPA at the input SC sampling network.

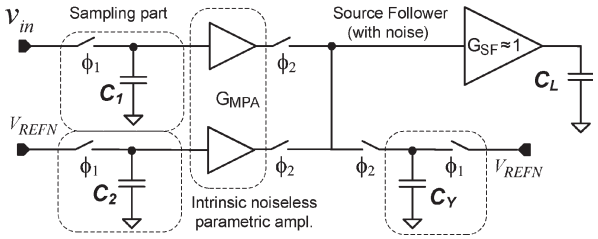


Fig. 10. Simplified MDAC model for noise analysis.

V. Noise Analysis

The pipeline structure of the ADC dictates that the noise budget for each MDAC stage has to be weighted by the total gain of precedent stages when referred to the converter's input. Although MPA is intrinsically noiseless, the different kT/C components (where k is the Boltzmann's constant, and T is the absolute temperature) have to be taken into account. Additionally, other noise sources increase the total noise power at the SFs' outputs during the amplification phase, including substrate noise and noise from the active devices used in the SF circuits [4].

Since the digital noise-coupling contribution through the substrate can be reduced by using well-known layout techniques, it can be neglected in this analysis. The total ADC noise performance is mostly dependent on the front-end S/H and on the first and second 1.5-bit MDACs. Hence, a simplified MDAC noise model, as shown in Fig. 10, can be used for noise analysis since the S/H can be treated as an MDAC in the Y configuration.

The input-referred noise of each MDAC can be described by

$$\overline{v_{i,n,\mathrm{MDAC}}^2} = 2 \cdot \left[\frac{kT}{C_1} + \frac{kT}{C_2} + \frac{kT}{C_Y} \cdot \frac{1}{(G_{\mathrm{MPA},C1})^2} \right] + 2 \cdot \left[\left(\frac{kT \cdot \gamma \cdot E_{\mathrm{SF}}}{C_L} \right) \cdot \left(\frac{1}{(G_{\mathrm{MPA},C1} \cdot G_{\mathrm{SF}})^2} \right) \right]. \quad (2)$$

The first term represents the different kT/C noise contributions related to the parametric and constant MOS capacitances (biased in the inversion region), and G_{MPA} represents the intrinsic noiseless gain of the parametric structure [4]. The remaining terms describe the noise from the SF at the output node, which has to be input referred. G_{SF} is the gain of the SF, which is slightly lower than unity mainly due to the body effect, E_{SF} is the SF excess thermal noise factor ($E_{\mathrm{SF}} \cong 2$), and γ is the transistor's excess noise factor ($\gamma \cong 1$). Since each stage is connected in a differential schematic, a factor of 2 is also affecting (2).

Considering that all MDACs have a similar input-referred noise level, the overall ADC input noise power is given by

$$\overline{v_{i,n,\mathrm{ADC}}^2} \approx \frac{v_{i,n,\mathrm{MDAC}}^2}{1} + \sum_{j=1}^{6} \frac{v_{i,n,\mathrm{MDAC}}^2}{4^j}. \quad (3)$$

Additional noise components, such as quantization noise, RMS jitter noise, and differential nonlinearity (DNL) "grass" noise, increase the total input-referred noise power of the ADC. For an 8-bit ADC and considering a full-scale input signal, a noise power term due to 2-ps RMS jitter (corresponding to a 35.5-μVRMS noise voltage) and a DNL "grass" at 1/3 least significant bit (LSB) level (corresponding to 230 μVRMS), the expected SNR is about 40.5 dB (assuming $C_1 = 312$ fF, $C_2 = 51$ fF, $C_Y = 25$ fF, and $C_L = 0.5$ pF).

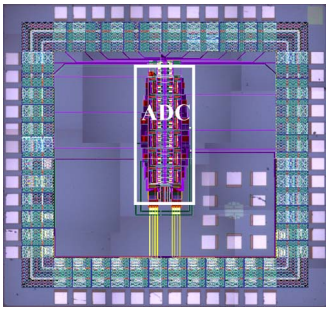


Fig. 11. Chip die photo (with overlaid layout plot). Die area ≈ 0.12 mm^2.

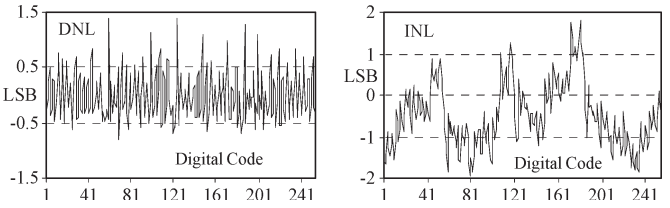


Fig. 12. DNL and INL plots at a conversion rate of 120 MS/s.

VI. Implementation and Measurement Results

An ADC prototype IC (micrograph shown in Fig. 11) was fabricated in a 130-nm 1P 8M CMOS pure logic process. To reduce mismatch effects, all MPA units in a stage of both channels have been laid out together using a common-centroid approach complemented by intensive postlayout simulations.

The ADC features an active area below 0.12 mm^2 and dissipates less than 14 mW at 120 MS/s and a 1.2-V supply. This total power, which excludes the external band-gap references, is distributed through digital correction (29%), the MPA cells and SFs (34%), and the comparator's PFDL and clock buffers (37%). Notice that this ADC was not power optimized since all stages are equally sized to minimize layout effort. Since the SNR of a pipeline ADC is mainly determined by the front-end S/H, first and second stages [10], by downscaling the others, a power savings up to 50% could be reached in each MPA unit. From [11], the power-and-area figure-of-merit can be calculated with FOM $= P \cdot A/(2^{\mathrm{ENOB}} \cdot F_S)$, where P and A are the total power and

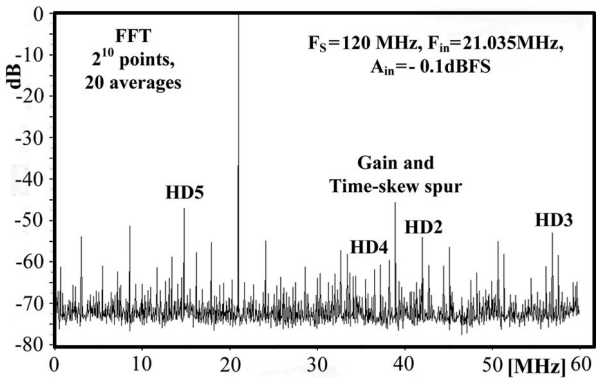


Fig. 13. Measured FFT results for $f_{\rm in} = 21$ MHz (at -0.1 dBFS).

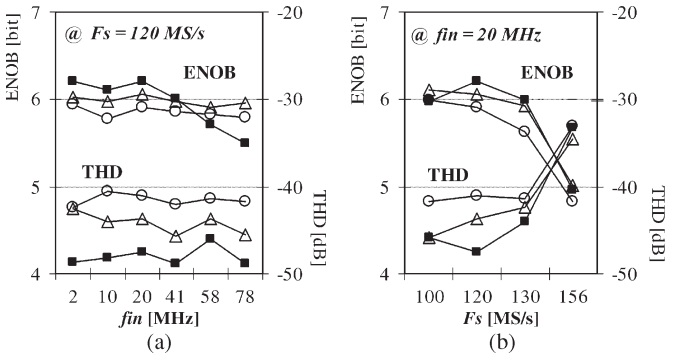


Fig. 14. Measured ENOB and THD for several (a) $f_{\rm in}$ (at -0.1 dBFS and $F_s = 120$ MS/s), and (b) F_s (at -0.1 dBFS and $f_{\rm in} = 20$ MHz). Results from three different samples.

TABLE I
KEY FEATURES AND MEASURED RESULTS

Technology	CMOS 130 nm 1P 8M
Supply Voltage	1.2 V
Resolution, Sampling Rate (F_S)	8-bit, 120 MS/s
Input full-scale (differential)	200 mVpp
Internal signal swing (differential)	400 mVpp
Power Consumption, Die Area	14 mW, 0.12 mm^2
DNL ; INL	-0.8/+1.4 ; -2.0/+2.0 LSB@8bit
Meas. @F_S=120 MS/s , @ f_{in}=20 MHz , @-0.055 dBFS	
SNR, SFDR	39.7 dB, 49.3 dB
THD, SNDR	-47.5 dB, 39.1 dB
ENOB	6.2 bits

area, respectively. A FOM better than 191 fJ $*$ mm^2/conversion was achieved with this ADC. All capacitors are implemented only using MOS devices. Fig. 12 shows that the measured DNL and integral nonlinearity (INL) errors at 120 MS/s are within $-0.8/+1.4$ LSB and ± 2.0 LSB, respectively. Fig. 13 displays a measured fast Fourier transform (FFT) for a 21-MHz input signal frequency $f_{\rm in}$ and a 120-MS/s sampling frequency F_S. The circuit achieves a peak SNR of 39.7 dB, a spurious-free dynamic range (SFDR) of 49.3 dB, and a peak total harmonic distortion (THD) of -47.5 dB, corresponding to an effective number of bits (ENOB) of 6.2 bits. The dynamic performance of this ADC was experimentally evaluated in three different samples (mounted using direct bonding) by sweeping the input signal frequency at $F_S = 120$ MS/s, as shown in Fig. 14(a), and by a sweeping the sampling rate for $f_{\rm in} = 20$ MHz, as shown in Fig. 14(b). For $F_S = 120$ Hz, the ADC exhibits a flat ENOB higher than 6 bits (with a peak of 6.2 bits) up to $f_{\rm in} = 41$ MHz [Fig. 14(a)]. The same flatness behavior in the ENOB is achieved for sampling rates up to 130 MS/s [Fig. 14(b)]. The ENOB variation between channels ranges from 0.1 bit for the best sample to 0.4 bit for the worst sample. Key features and measurement results are summarized in Table I.

VII. CONCLUSION

This brief has presented a MOS only 8-bit 120-MS/s interleaved pipeline ADC that extensively uses DT MPA. Measurements for a 20-MHz input signal shows that the ADC achieved, without calibration, 39.7 dB of SNR, 49.3 dB of SFDR, -47.5 dB of THD, 39.1 dB of SNDR, and 6.2 bits of ENOB.

ACKNOWLEDGMENT

The authors would like to thank their colleagues B. Esperança, J. Taffarel, J. Paisana, and N. Paulino for their contributions and Prof. Y. Tsividis for his initial feedback and motivation.

REFERENCES

[1] B. Murmann and B. E. Boser, "A 12-bit 75-MS/s pipeline ADC using open-loop residue amplification," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2040–2050, Dec. 2003.

[2] J. K. Fiorenza, T. Sepke, P. Holloway, C. G. Sodini, and H.-S. Lee, "Comparator-based switched-capacitor circuits for scaled CMOS technologies," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2658–2668, Dec. 2006.

[3] L. Brooks and H.-S. Lee, "A zero-crossing-based 8-bit 200 MS/s pipeline ADC," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2677–2687, Dec. 2007.

[4] S. Ranganathan and Y. Tsividis, "Discrete-time parametric amplification based on a three-terminal MOS varactor: Analysis and experimental results," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2087–2093, Dec. 2003.

[5] J. Hu, N. Dolev, and B. Murmann, "A 9.4-bit, 50-MS/s, 1.44-mW pipelined ADC using dynamic source follower residue amplification," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1057–1066, Apr. 2009.

[6] A. Yoshizawa and S. Iida, "A gain-boosted discrete-time charge-domain FIR LPF with double-complementary MOS parametric amplifiers," in *Proc. IEEE ISSCC*, 2008, pp. 68–69.

[7] S. Magierowski, H. Chan, and T. Zourntos, "Subharmonically pumped RF CMOS paramps," *IEEE Trans. Electron Devices*, vol. 55, no. 2, pp. 601–608, Feb. 2008.

[8] J. P. Oliveira, J. Goes, N. Paulino, J. Fernandes, and J. Paisana, "New low-power 1.5-bit time-interleaved MDAC based on MOS capacitor amplification," in *Proc. IEEE ICECS*, 2008, pp. 251–254.

[9] J. P. Oliveira, J. Goes, B. Esperança, N. Paulino, and J. Fernandes, "Low-power CMOS comparator with embedded amplification for ultra-high-speed ADCs," in *Proc. ISCAS*, 2007, pp. 3602–3605.

[10] J. Goes, J. C. Vital, and J. E. Franca, "Systematic design for optimization of high-speed self-calibrated pipelined A/D converters," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 12, pp. 1513–1526, Dec. 1998.

[11] J. Sauerbrey, T. Tille, D. Schmitt-Landsiedel, and R. Thewes, "A 0.7-V MOSFET-only switched-opamp $\Sigma\Delta$ modulator in standard digital CMOS technology," *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1662–1669, Dec. 2002.

C.2 MIXDES'09

LNA, Oscillator and Mixer Co-Design for Compact RF-CMOS ISM Receivers

MIXDES 2009, 16th International Conference

LNA, Oscillator and Mixer Co-Design for Compact RF-CMOS ISM Receivers

J. Ferreira, I. Bastos, L. Oliveira, J.P. Oliveira
CTS - Centre of Technology and Systems
UNINOVA / FCT / UNL
Monte da Caparica, Portugal
jpao@fct.unl.pt

T. Michalak[1], P. Pankiewicz[1], B. Nowacki[1], P. Makosa[1], A. Rybarczyk
Chair of Computer Engineering
Poznan University of Technology
Poznan, Poland

Abstract—**A co-design strategy for the implementation of a low-voltage fully integrated CMOS receiver is presented. This co-design approach allows the design of a compact direct-conversion receiver by avoiding 50 Ω matching buffers and networks, and AC coupling capacitors between mixer inputs and LNA and oscillator outputs. Moreover, the proposed circuit does not require DC choke inductors for mixer biasing. Since a 1.2 V power supply is used, a current bleeding technique is applied in the LNA and in the mixer. To avoid inductors and obtain differential quadrature outputs, an RC two-integrator oscillator is employed, in which, a filtering technique is applied to reduce phase noise and distortion. The proposed receiver is designed and simulated in a 130 nm standard CMOS technology. The overall conversion voltage gain is higher than 35.8 dB and the noise figure is 5.3 dB.**

Index Terms—**LNA, RC Oscillator, mixer, direct-conversion receiver, low–IF receiver, fully integrated CMOS receiver.**

I. Introduction

The direct conversion architecture, shown in Fig. 1, and low-IF architecture, shown in Fig. 2 are approaches to enable full integration of RF receivers in pure standard digital CMOS technologies, which by their turn are reaching higher transistor's cutoff frequencies ω_T. The success of these approaches is supported by its dissemination from high demanding 2G and 3G handsets to low data rate and low-power wireless sensors (WSN) and ISM applications.

Both, the direct conversion receiver (DCR) and the low-IF receiver techniques, allow significant reduction of the number of off-chip components, which means that all the major building blocks will interconnect to each other inside the chip [1-3]. Therefore, the match between these internal interconnects at 50 Ω level is no longer required. This simple approach proposed here permits a highly integrated, low area, low power, and low-cost implementation.

DCR and low-IF receivers require linear low noise amplifier (LNA) and a mixer with a high frequency local oscillator (LO) with precise quadrature outputs. In these types of receivers, the conventional approach of designing independently these blocks is not longer suitable. Alternatively, a co-design methodology for adapting the mixer to the LNA and to the oscillator is required. All these requirements are difficult to fulfill simultaneously, and therefore, an optimized trade-off process should be followed.

This paper proposes a co-design strategy applied jointly to the LNA, mixer and to the local oscillator for applications in the sub-gigahertz ISM-band and with low to moderate data rate, which can be applied to direct or low-IF receivers.

The main objective of this co-design is to avoid matching buffers (in LNA and oscillator outputs), and directly connect to the mixer without using AC coupling capacitors and choke inductors. This co-design also allows the minimization of power consumption, which is an ongoing research work.

II. RF Front-End Considerations

In the homodyne receiver shown in Fig. 1, the RF spectrum is directly translated to the baseband in a single downconversion step. This step is performed by the complementary operation of the LNA, mixer and LO. The sub-gigahertz RF signal is first amplified by the LNA and then down-converted to zero-IF in-phase and quadrature (I/Q) signals by the composite mixer driven by quadrature LO signals. Since the LO has the same frequency of the input radio carrier, in DCR the I/Q signals are needed to separate the wanted channel from its mirror, which is accomplished by means of a Hilbert transform. Therefore, this downconversion requires accurate quadrature signals generated by the local oscillator [1, 2].

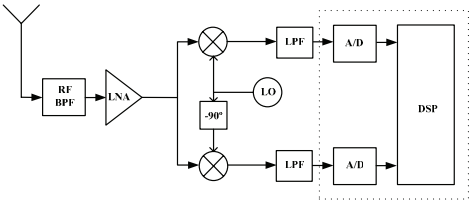


Figure 1. Direct Conversion Receiver

[1] MSc. students at FCT- Universidade Nova de Lisboa, Portugal, under ERASMUS interchange protocol.

All the remaining mixed-mode processing is performed at the baseband, relaxing the requirements for filters and A/Ds.

Direct conversion receivers (DCR) have several design issues to be addressed, which are related to flicker noise, LO (local oscillator) leakage, quadrature errors, DC offsets:

a) Flicker noise – Having a 1/f corner at low frequency, this noise can corrupt substantially the low frequency baseband signals, which is a severe problem in MOS implementation. The major contribution at the output comes from the current commutating switching transistors of the mixer.

b) LO leakage – LO signal coupled to the antenna will be radiated again and re-injected to the mixer through the main signal path, originating unwanted baseband DC components.

c) Quadrature error – Quadrature error and mismatches between the amplitudes of the I and Q signals corrupt the downconverted signal constellation.

d) DC offsets – Since the downconverted band extends down to zero frequency, any offset voltage can corrupt the signal and saturate the receiver's baseband output stages. Hence, DC offset removal or cancellation is required in direct-conversion receivers.

The DCR approach removes the IF high-Q filters (reducing the receiver area and/or avoiding external components) and therefore the LNA can be directly connected to the mixer. Since the input mixer impedance is essentially capacitive the LNA output does not have to be matched to 50 Ω. For a gate input type mixer, meaning that it is driven by a voltage, it is the LNA voltage gain that should be considered.

It is well known that heterodyne receivers have important limitations due to the use of external image reject filters. DCR receivers have some drawbacks because the signal is translated directly to the baseband. Thus, there is interest in the development of new techniques to reject the image without using filters. An architecture, which combines the advantages of both the IF and the zero-IF receivers, is the low-IF architecture.

The low-IF receiver is a heterodyne receiver that uses special mixing circuits that cancel the image frequency, as shown in Fig. 2. A high quality image reject filter is not necessary anymore, while the disadvantages of the zero-IF receiver are avoided [3, 4].

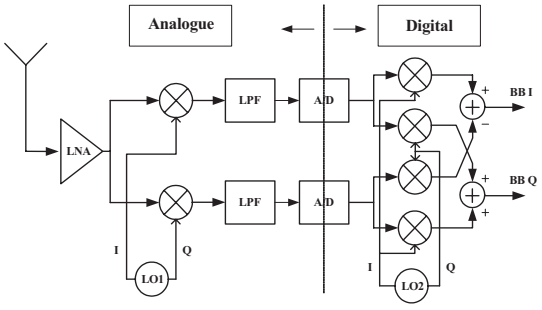


Figure 2. Low-IF receiver (simplified block diagram).

In this paper, we present a co-design strategy that can be applied for DCR or low-IF receivers.

III. RF Front-End Key Blocks

A. LNA

The LNA, shown in Fig. 3, uses the source-degenerated topology around input transconductance transistor M_1. This architecture is very common among narrowband LNA's as it is very close to achieving the goal of providing the input match and best noise performance simultaneously [1-4]. The cascode transistor (M_2) is used to reduce the effect of the gate-drain capacitance C_{gd} of the input transistor (M_1) and to increase the reverse isolation of the LNA. This improves the stability and makes the LNA's input impedance less sensitive to its load impedance. The number of integrated inductors is reduced to one, since L_S is implemented with the bonding wire and the output inductance L_D is replaced by a resistance. In order to avoid significant voltage drop at the output resistor a bleeding current is injected at the drain of M_1 preserving the value of the g_m of M_1, needed to maintain the input matching to 50 Ω.

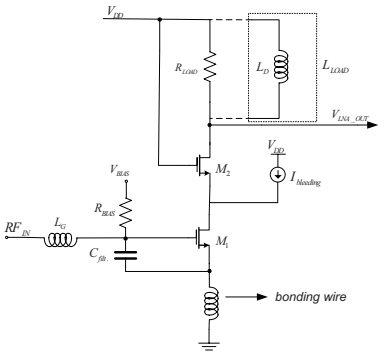


Figure 3. LNA circuit schematic.

The input impedance of the LNA is approximately given by:

$$Z_{in} = j\omega L_g + j\omega L_s + \frac{1}{j\omega(C_{filt} + C_{gs1})} + \frac{g_{m1} L_s}{C_{filt} + C_{gs1}} \quad (1)$$

On one hand, it is clear from (1) that at resonance and for a given L_S the 50 Ω input match sets the value of the transconductance gain. On the other hand, due to capacitive nature of the mixer input, an optimized value for the LNA voltage gain can be found by selecting an appropriate value of the output LNA load resistance.

Entering into account the effective transconductance gain due to the input matching requirement, the LNA voltage gain in case of resistive load R_L is given approximately by:

$$|A_v| \approx \frac{R_L}{2\omega_0 L_s} \quad (2)$$

The value of R_L is the result of co-designing the LNA together with the mixer. The value of the 700 Ω guaranteed the best performance, from which a gain of 28 dB is achievable.

B. Quadrature Local-Oscillator

The schematic of the two-integrator oscillator [5] is presented in Fig. 4. Each integrator is realized by a differential pair (transistors M) and a capacitor (C). The oscillator frequency is controlled by I_{tune}. There is an additional differential pair (transistors M_L), with the output cross-coupled to the inputs, which performs two related functions: a) compensation of the losses due to R to make the oscillation possible (a negative resistance is created in parallel with C); b) amplitude stabilization, due to the non-linearity (the current source I_{level} controls the amplitude). To start the oscillations the condition $g_m > 1/R$ must be met. Moreover, the I_{level} is used to control the output signals amplitude.

In order to obtain low distortion output, a filtering technique is used. To achieve this goal the extra capacitor C_{filter} is introduced to the terminals of the tuning current source I_{tune}. The introduction of this element reduced cancels the harmonics at this point and reduced the oscillator phase-noise.

The circuit of Fig. 4 can be represented by the linear model in Fig. 5, where the negative resistance is realized by the cross-coupled differential pair (M_L), and R represents the integrator losses due to the pairs of resistances R / 2.

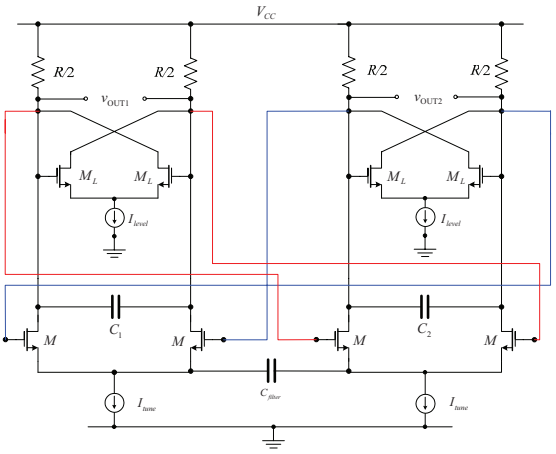


Figure 4. Two-integrator schematic.

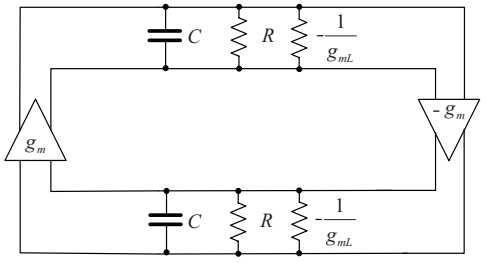


Figure 5. Two-integrator linear model.

The oscillator frequency varies by changing either the capacitance or the transconductance. In a practical circuit we can use varactors to change the capacitance or, most commonly, we can change the tuning current, and therefore, the transconductance.

These oscillators have wide tuning range with very precise inherent differential quadrature outputs (less than one degree quadrature error), which are required for very compact DCR and low-IF receivers [5].

C. Mixer

The I/Q mixer topology, shown in Fig. 6, is based in a active double balanced Gilbert cell. The double-balanced mixer is a complex circuit, which has I/Q LO differential inputs. This mixer has higher gain, lower noise figure, good linearity, high port-to-port isolation, high spurious rejection, and less even-order distortion, with respect to the single-balanced mixer. The main disadvantage is the increased area (due to complexity) and power consumption; additionally, in order to save area and since LNA output is single ended we have not used a balun transformer to provide the RF differential at the mixer input [1-4].

The main objective of this work is the co-design between the LNA, LO and mixer in order to avoid 50 Ω matching buffers and AC coupling capacitors.

Considering the high impedance mixer input, the LNA output can be directly AC connected to the mixer. Nevertheless, the LNA output DC component is important to bias the transconductance mixer stage, which controls the mixer conversion gain. In this transconductance mixer block, a minimum L is used, to maximize gain and speed, and the W is adjusted according DC LNA output voltage. An additional current is injected into the mixer transconductance (formed by M_3 and M_4) to improve linearity and setting the conversion gain and noise figure. By adjusting this current, the DC output level from the LNA will have less impact on the output voltage.

The mixing switching current commutating stage is formed by NMOS pair transistors, which are connected directly to the oscillator I/Q outputs. As in the previous stage, the oscillator AC output is connected directly to the mixer. The oscillator amplitude needs to be maximized to properly drive the mixer switching transistors and reduce the mixer output $1/f$ noise. Moreover, the oscillator output DC component is important to bias these switching pairs, and it will define their widths (since the L is kept at minimum value).

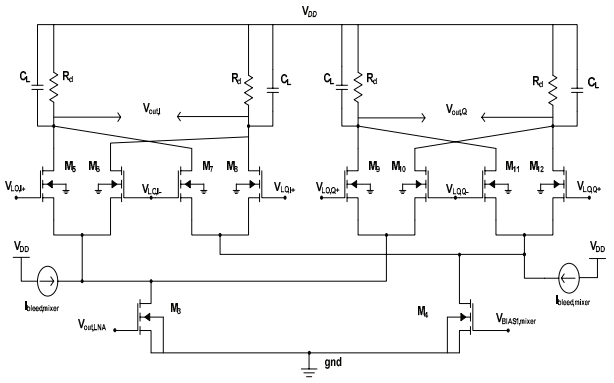


Figure 6. Mixer circuit schematic.

D. Co-design strategy

The co-design of the LNA, oscillator and mixer facilitates the optimization process to reach better tradeoff between conversion gain and noise figure. LNA and mixer merged topologies has been proposed in literature [6], but in our design a cascaded structure was chosen. This topology can achieve higher gain and better noise figure since the noise contribution of the mixer can be substantially suppressed by the high voltage gain of the LNA. In this work we propose that the mixer should be co-designed with both the LNA and LO.

The design process begins by maximizing the LNA voltage gain for a given input match criteria (50 Ω in this case). By its turn the oscillator is designed to maximize the signal output swing voltage and improve the I/Q signals accuracy for a given power. With the obtained DC components at the output of these blocks, the mixer is then optimized to reach a reasonable conversion gain and noise figure. In order to reduce the total area, the design must remove as much as possible the use of inductors and AC coupling capacitors.

The co-design example presented in this work will reduce the total circuit area allowing the design of a low cost compact receiver.

IV. Simulation Results

To validate the proposed strategy, the LNA, mixer and quadrature oscillator has been co-designed and simulated on a 130 nm CMOS technology with 1.2 V power supply.

From the traditional inductive load LNA (we used a 27nH inductor, one of the maximum available value from the chosen technology), with a 50 Ω output we obtained a 15 dB of gain and 2.7 dB of noise figure at 900 MHz. Replacing this inductor by a 700 Ω resistor (which can not be higher due to power supply headroom) we obtain a 28 dB voltage gain with a 2.24 dB noise figure. The power consumption is lower than 9 mW for the two cases.

Concerning the RC two-integrator oscillator, we have used a capacitive filtering technique to reduce phase noise and distortion. These results of the performed analysis are presented in table 1.

TABLE I. RC Oscillator results

Case	*Phase noise @ 10MHz offset* [dBc/Hz]	*THD* [dB]	P_D [mw]
I: without C_{filter}	-111.0	-31.3	4.8
II: with C_{filter}	-112.6	-30.8	5.2
III: with C_{filter} optimized for THD	-120	-43.5	8.6

For cases I and II the simulation results show that the output of the circuit with filtering capacitor C_{filter} has better value of phase noise (1.6 dBc/Hz lower). Introducing this capacitor causes a drop of magnitude. Due to this fact, in order to achieve the assumed amplitude and frequency level, the current values have to be increased. Therefore, the power consumption slightly rises. In case III higher I_{tune} current was applied and the impact of filtering capacitor C_{filter} is reduced because it doesn't change so significantly the values of phase noise and THD.

The final parameters of the designed oscillator are following: R = 314 Ω, (W/L) = 15 µm / 0.255 µm for M transistors, and W/L = 10.8 µm / 0.255 µm for ML transistors, C_{filter} =217 fF (for cases I and II) and C_{filter} =430 fF (for case III), C_{filter} =202 fF. The oscillator output differential amplitude is 290 mV @ 900 MHz.

The final parameters of the designed mixer are the following: R_L=800 Ω, C_L=2.5 pF, (W/L)= 100 µm/ 0.13 µm for the switching stage transistors, (W/L)= 30 µm / 0.13 µm for the RF stage transistors and $I_{bled,mixer}$=4.15 mA.

In Figs. 7 to 9 are presented the simulations results for the completed front-end obtained from SpectreRF simulator, using BSIM3V3 models, including noise.

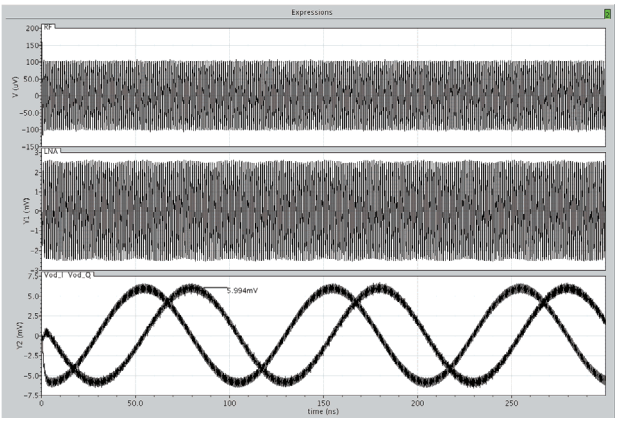


Figure 7. Periodic time response.

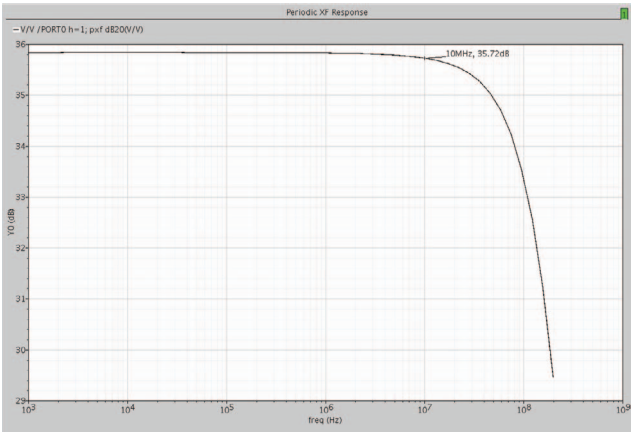


Figure 8. Total conversion gain.

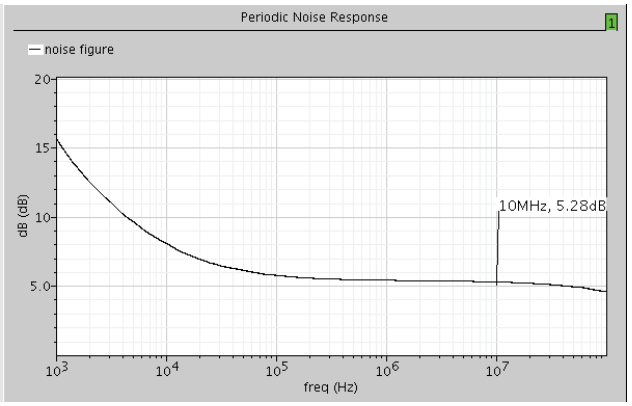


Figure 9. Cascade noise figure.

V. Discussion and Conclusions

In this paper, a co-design strategy for the implementation of a low-voltage, low-area, low-cost, fully integrated CMOS receiver was presented. This approach avoids 50 Ω matching buffers and networks, AC capacitors coupling, and DC choke inductors.

We present a resistive load LNA, with 700 Ω load and inductor-less differential RC quadrature oscillator, which are combined with a mixer in a co-design strategy. A current bleeding technique was applied at the LNA and mixer, due to the low power supply voltage. The low area quadrature two-integrator oscillator uses a capacitive filtering technique, which reduces the oscillator phase-noise and the harmonic distortion.

The approach presented in this paper has only one inductor allowing the design of a very compact and low cost receiver (DCR or low-IF), which is required for low data rates ISM applications.

The proposed receiver was designed and simulated in UMC 130nm CMOS technology. The total conversion voltage gain is 35.8 dB and the cascade noise factor is 5.3 dB for the interest band.

Acknowledgment

This work was supported by the Portuguese Foundation for Science and Technology (FCT) through funding of CTS-UNINOVA and project LEADER (PTDC/EEA-ELC/69791/2006).

References

[1] B. Razavi, *RF Microelectronics*, Prentice-Hall, 1998.

[2] T. H. Lee, The Design of CMOS Radio Frequency Integrated Circuits (2nd edition), Cambridge University Press, 2004.

[3] J. Crols and M. Steyaert, "A Single-Chip 900 MHz CMOS Receiver Front-end with High Performance Low-IF Topology", *IEEE J. Solid-State Circuits,* vol. 30, nº 12, pp. 1483–1492, December 1995.

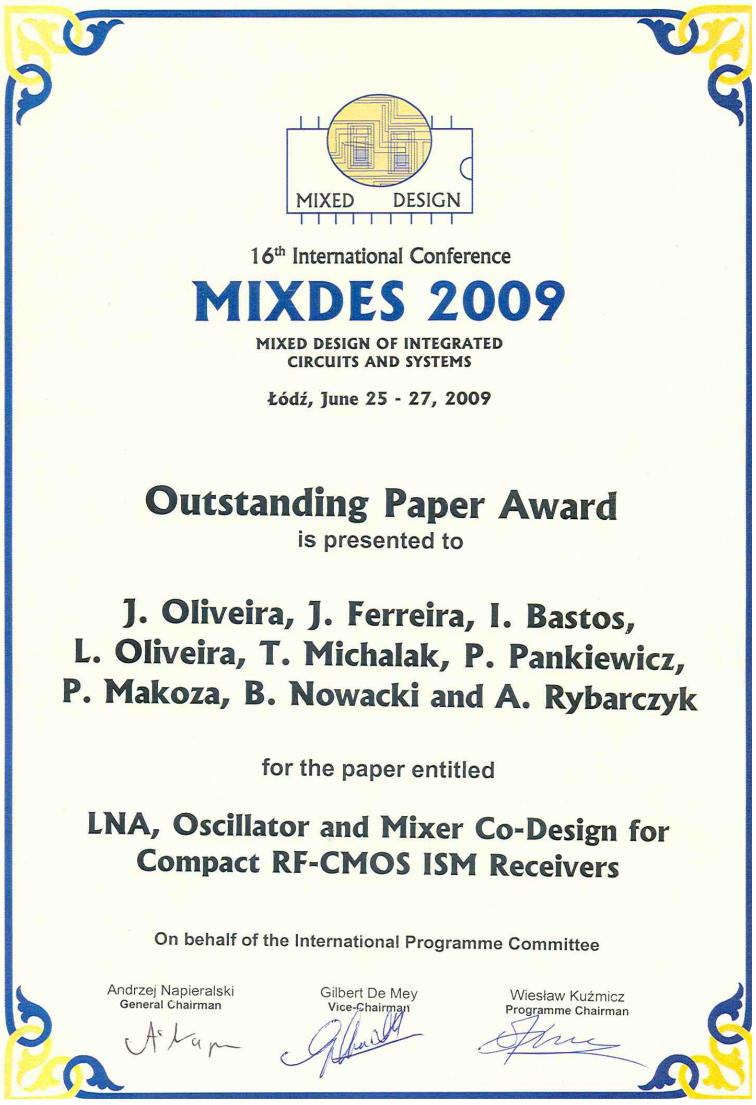
[4] J. Crols and M. Steyaert, *CMOS Wireless Transceiver Design*, Kluwer, 1997.

[5] L. Oliveira, J. Fernandes, I. Filanovsky, C. Verhoeven, M. Silva, "*Analysis and Design of Quadrature Oscillators*" Springer: Analog Circuits and Signal Processing 2008

[6] H. Sjoland A. Karimi, and A. A. Abidi, "A Merged CMOS LNA and Mixer Fro a WCDMA Recevier," IEEE J. Solid-State Circuits, vol. 38, no. 6, pp. 1045-1050, 2003.

Appendix D

Award


MIXED DESIGN

16th International Conference

MIXDES 2009

MIXED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS

Łódź, June 25 - 27, 2009

Outstanding Paper Award

is presented to

J. Oliveira, J. Ferreira, I. Bastos, L. Oliveira, T. Michalak, P. Pankiewicz, P. Makoza, B. Nowacki and A. Rybarczyk

for the paper entitled

LNA, Oscillator and Mixer Co-Design for Compact RF-CMOS ISM Receivers

On behalf of the International Programme Committee

Andrzej Napieralski, General Chairman

Gilbert De Mey, Vice-Chairman

Wiesław Kuźmicz, Programme Chairman

Bibliography

[1] J. Oliveira, J. Goes, M. Figueiredo, E. Santin, J. Fernandes, and J. Ferreira, "An 8-bit 120-ms/s interleaved cmos pipeline adc based on mos parametric amplification," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 57, pp. 105 –109, feb. 2010.

[2] J. Ferreira, I. Bastos, L. Oliveira, J. Oliveira, T. Michalak, P. Pankiewicz, B. Nowacki, P. Makosa, and A. Rybarczyk, "Lna, oscillator and mixer co-design for compact rf-cmos ism receivers," *Mixed Design of Integrated Circuits Systems, 2009. MIXDES '09. MIXDES-16th International Conference*, pp. 291 –295, june 2009.

[3] J. Oliveira, L. Oliveira, J. Ferreira, I. Bastos, T. Michalak, P. Pankiewicz, B. Nowacki, P. Makosa, and A. Rybarczyk, "Co-design strategy approach of lna, oscillator, and mixer," *International Journal of Electronics And Telecommunications Quarterly*, vol. 55, no. 3, pp. 37 –51, 2009.

[4] A. Abidi, "Low-power radio-frequency ics for portable communications," *Proceedings of the IEEE*, vol. 83, pp. 544–569, Apr 1995.

[5] B. Razavi, *RF Microelectronics*. Prentice Hall, 1998.

[6] L. B. Oliveira, J. R. Fernandes, I. M. Filanovsky, C. J. Verhoeven, and M. M. Silva, *Analysis and Design of Quadrature Oscillators*. Springer, 2008.

[7] B. Razavi, "Design considerations for direct-conversion receivers," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 44, pp. 428–435, Jun 1997.

[8] H.-M. Seo, Y. Moon, Y.-K. Park, D. Kim, D.-S. Kim, Y.-S. Lee, K.-H. Won, S.-D. Kim, and P. Choi, "A low power fully cmos integrated rf transceiver ic for wireless sensor networks," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 15, pp. 227 –231, feb. 2007.

[9] S. Sarkar, P. Sen, A. Raghavan, S. Chakarborty, and J. Laskar, "Development of 2.4 ghz rf transceiver front-end chipset in 0.25μm cmos," *VLSI Design, 16th International Conference on*, pp. 42 – 47, jan. 2003.

[10] M. Terrovitis and R. Meyer, "Noise in current-commutating cmos mixers," *Solid-State Circuits, IEEE Journal of*, vol. 34, pp. 772 –783, jun 1999.

[11] F. T. S. Douss and M. Loulou, "Design optimization methodology of cmos active mixers for multi-standard receivers," *International Journal of Electronics, Circuits and Systems*, vol. 1, no. 1, 2007.

[12] http://www.cadence.com.

[13] http://www-device.eecs.berkeley.edu/ bsim3/.

[14] K. M. David A. Johns, *Analog Integrated Circuit Design.* Wiley, 1997.

[15] F. Maloberti, *Data Converters.* Springer, 2008.

[16] M. S. B. Hamid Movahedian, Behnam Sedighi, "Wide-range single-ended cmos track-and-hold circuit," *IEICE Electron*, vol. 4, no. 12, pp. 400 – 405, 2007.

[17] T. Y. Ranganathan, S., "Discrete-time parametric amplification based on a three-terminal mos varactor: analysis and experimental results," *Solid-State Circuits, IEEE Journal of*, vol. 38, pp. 2087 – 2093, Dec. 2003.

[18] P. Figueiredo and J. Vital, "The mos capacitor amplifier," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 51, pp. 111 – 115, march 2004.

[19] B. McCarroll, C. Sodini, and H.-S. Lee, "A high-speed cmos comparator for use in an adc," *Solid-State Circuits, IEEE Journal of*, vol. 23, pp. 159–165, Feb 1988.

[20] J. Juan-Chico, M. J. Bellido, A. J. Acosta, M. Valencia, and J. L. Huertas, "Analysis of metastable operation in a cmos dynamic d-latch," *Analog Integr. Circuits Signal Process.*, vol. 14, no. 1-2, pp. 143–157, 1997.

[21] P. Cusinato, M. Bruccoleri, D. Caviglia, and M. Valle, "Analysis of the behavior of a dynamic latch comparator," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 45, pp. 294–298, Mar 1998.

[22] J. Oliveira, J. Goes, B. Esperanca, N. Paulino, and J. Fernandes, "Low-power cmos comparator with embedded amplification for ultra-high-speed adcs," *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, pp. 3602 –3605, may 2007.

[23] J. Fernandes and M. Silva, "Asymmetric latches and nonideal performance of parallel adcs," *Electrotechnical Conference, 1994. Proceedings., 7th Mediterranean*, pp. 629 –632 vol.2, apr 1994.

[24] MAXIM, application note 1040: http://www.maxim-ic.com.